Host-Microbiota Interactions

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Ce Yuan

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Advisors:

Subbaya Subramanian, M.S., Ph.D.

Walter C. Low, Ph.D.

July 2022

Acknowledgments

It has been a long and bumpy ride.

My graduate school years are no doubt the most unforgettable time of my life. The first 2 years of my Ph.D. life threw me deep into depression, which ensnared me for the better part of the past 7 years.

I was grateful to have met with Dr. Subbaya Subramanian in 2016, whose gentle encouragement, mentorship, and belief in my crazy ideas helped me pick myself back up. It was then I made perhaps one of the most difficult decisions a Ph.D. student can make, changing major and mentor. It is from that time, I have found my passion and drive, which helped me get to where I am today. I am forever grateful for Dr. Subramanian's mentorship and more importantly friendship.

During my time at Dr. Subramanian's laboratory, he always encouraged me to seek and experiment with my own development goals and career interests. That is why I spent many hours outside of his laboratory exploring different things and honing my skills in bioinformatics, patent law, finance, and business consulting. During that time, I spent a few years at the Stem Cell Institute (SCI), which afforded me a place to validate myself, develop my skills, establish collaborations, and most importantly finding new mentorships and friendships. I am grateful for everyone at the SCI, especially Drs. Walter Low and Nobuaki Kikyo, who provided me with countless opportunities.

To all the members of Subree Lab, Lihua Li, Xianda Zhao, Dechen Wangmo, Travis Gates, Pearl Wilcox, and many more, who have helped me throughout my journey, I say thank you.

To everyone in the BICB program who gave me the opportunity and guided me throughout my time here, and those who served tirelessly on my committees, Drs. Claudia Neuhauser, Chad Myers, Yuk Sham, Michael Sadowsky, Ran Blekhman, Christopher Staley, Ajay Prakash, and many more, I say thank you.

To my parents, my in-laws, grandparents, my extended families, and all my friends around the world who were by my side and always believed in me, I say thank you.

I could not have done any of the work without generous support from the Healthy Food, Healthy Lives Institute, Minnesota Colorectal Cancer Research Foundation, and the University of Minnesota Informatics Institute. To all the donors, advocates, and administrators, I say thank you.

Last but definitely not least, I owe everything to my wife, Xiaoyu Sun, who has always believed in me, who has always supported me, and who brought us the best gift ever, our son Albert (b. 2019). To her, who *only occasionally* asked me "*When are you graduating?*", I say, now is the time.

Dedication

This dissertation is dedicated to my loving wife Xiaoyu Sun and our son Albert.

Abstract

Advances in high-throughput technologies in the past decade have helped delineate the important role of gut microbiota on human health. Changes in gut microbiota composition have consistently been observed in various health conditions, including colorectal cancer (CRC). Studies show that the microbiota fills in a variety of niche metabolic pathways that the host does not possess. For example, the microbiota produces butyrate, which provides the colon's epithelial cells with about 70% of their energy needs. The typically fast proliferation of tumor cells in CRC patients drastically alters the tumor's nutrient microenvironment. Those alterations correspond to changes in the microbiota composition and function. Yet, it is not entirely clear how the gut microbiota interacts with the host cells. In the present dissertation, we took a systems biology approach to map various interactions between the microbiota and the host. We presented comprehensive interaction networks between the tissue-associated microbiota, gene expression, metabolites, and immune cell infiltration.

First, we performed an integrated analysis between the mucosa-associated microbiota and the mucosa metabolome in healthy, nonhuman primates to investigate the metabolic interactions between the microbiota with its host. We found the microbiota composition was distinct at each tissue location, with variation by host individual also observed. Additionally, the microbiota-metabolome dynamics were primarily driven by interactions in the distal colon. More importantly, we found that the interaction network in the large intestines is sparse compared to

the network in the small intestines. We postulate that the sparse interaction is due to a specialized role of microbiota in the large intestines and may explain the associations between altered microbiota and CRC.

We then investigated the interaction between the microbiota and host in CRC tissues. Previous studies have shown that microRNAs (miRNAs) regulate gene expression and have important roles in cancer development and overall metabolism. In addition, recent studies suggested that host miRNAs can also regulate bacterial growth and influence the composition of the gut microbiome. Here, we investigated the link between the gut microbiota and the expression of miRNA in CRC. We found that dozens of miRNAs are differentially regulated in CRC tumors and adjacent normal colonic tissues and that these miRNAs are correlated with the abundance of microbes in the tumor microenvironment. Moreover, we found that microbes that have been previously associated with CRC were correlated with miRNAs that regulate genes related to interactions with microbes. Notably, these miRNAs likely regulate glycan production, which is important for the recruitment of pathogenic microbial taxa to the tumor. This work characterized a global relationship between microbial community composition and miRNA expression in human CRC tissues.

Finally, we investigated a network interaction between the microbiota, metabolites, and immune cell infiltrations in CRC tissues. We first found metabolic pathways related to essential amino acids enriched in the tumors. Further investigations through network analysis suggest that metabolites potentially

mediated the interactions between the microbiota and the host, including the immune cells.

Overall, using a multi-omics and systems biology approach, we provide comprehensive interaction networks of host-microbiota interactions. The network suggests a shortlist of interactions that contribute to a better understanding of the complex relationship and informs additional hypotheses for future validations and developing strategies for modulating microbiota in the intestines to improve host health.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Introduction

An average human intestine contains more than 100 trillion bacteria (collectively known as the gut microbiota)[1]. In recent decades, several studies have suggested that the gut microbiota is crucial to human health and the development of diseases[2–7].

In healthy humans, a key factor associated with microbiota variations is host genetics[8–11]. In a study of healthy twins, Goodrich et al. found that host genetics drive microbiota composition and can also affect the host metabolic phenotype[9]. Several other studies have found an association between the abundance of *Bifidobacterium* species and the presence of single-nucleotide polymorphisms (SNPs) in close proximity to the host lactase gene locus[8,12]. This association suggests that the *Bifidobacterium* species conceivably assist the host in metabolizing lactose.

Studies have also identified that altered microbiota composition and function (dysbiosis) is a common signature of CRC[2–7]. Bacterial candidates such as *Fusobacterium nucleatum* and *Bacteroides fragilis* are consistently enriched in tumor tissues and included in that signature. Specific factors in those bacteria, including FadA and Fap2 protein from *F. nucleatum* and *B. fragilis* toxins, have been identified that play a role in CRC pathobiology[13–21]. However, our knowledge of most other bacteria associated with the CRC microenvironment is limited. Moreover, we are just beginning to understand the complex interactions between host and microbiota in CRC, as well as other clinical disorders such as

neurodegenerative diseases[22]. A recent CRC study found that loss-of-function mutations in the mitogen-activated protein kinase (*MAPK*) and *Wnt* signaling pathways are associated with specific sets of microbiota profiles[2]. Furthermore, mutations in the tumor-suppressor adenomatous polyposis coli (*APC*) gene are also associated with a distinct inter-microbiota association network[2]. These findings suggest that a common factor might orchestrate the dynamic host-microbiota interaction(s) and functional relationship(s).

Based on mounting evidence, we postulate that the altered nutrient composition and gene expression in the CRC microenvironment selectively influences the surrounding microbiota, leading to alterations in its composition. In this dissertation work, we present our investigation of the role of microRNA and metabolites in mediating host-microbiota interactions in CRC **(Figure 1.1)**. After outlining the evidence pointing to their role, we reflect on the future direction of this rapidly evolving field.

1.2 Microbiota and Colorectal Cancer

Studies have consistently found a change in the microbiota compositions in CRC patients[6,23]. Intriguingly, several studies have shown that genetic CRC models and chemical-induced models develop significantly fewer CRC tumors in germ-free or antibiotic-treated animals[24–26]. These observations suggest a potentially robust effect of the intestinal microbiota in CRC development. In the healthy intestinal tract, the microbiota is dominated by the Bacteroidetes and Firmicutes phyla, which together comprise about 70% of the microbiota[23].

Several taxa of bacteria have been implicated in the microbiota of CRC patients.

In their stool samples, at the species level, a consistently higher abundance of

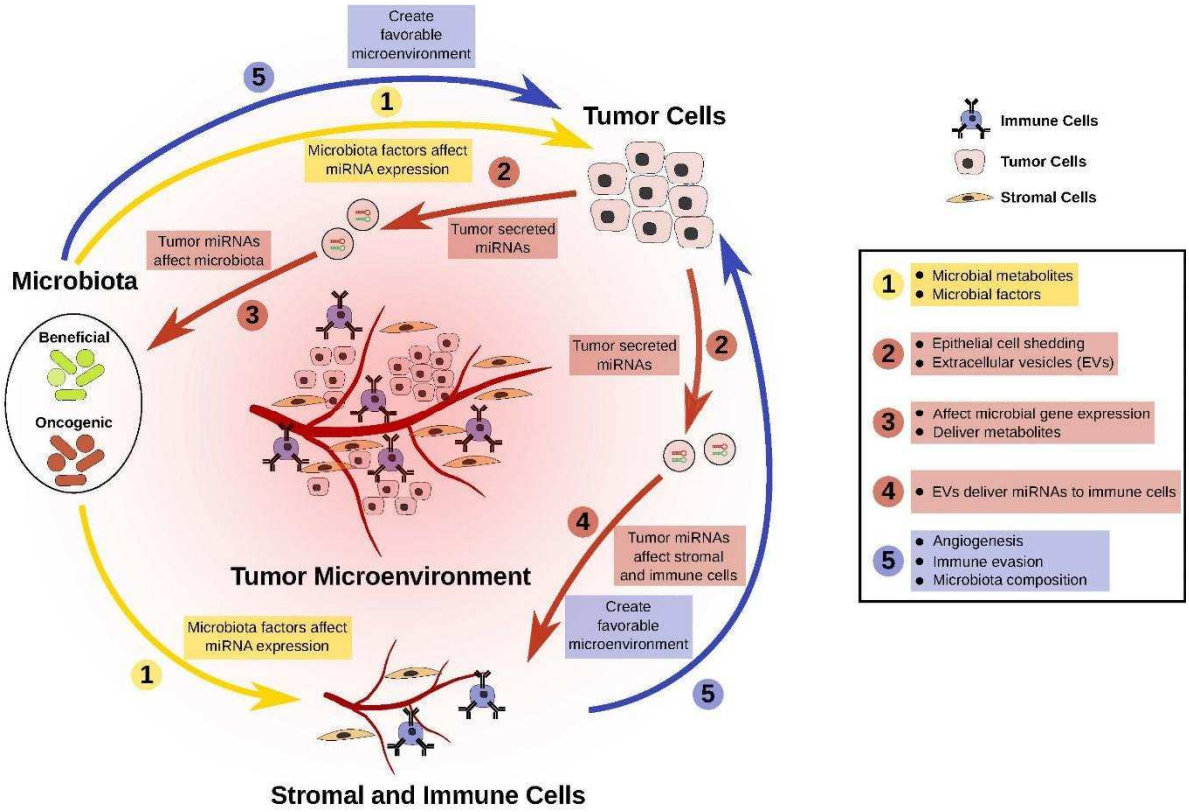

Figure 1.1. Host–microbiota interactions in colorectal cancer.
Microbiota composition has a functional effect on the cancer cells, via stromal and tumor-infiltrating immune cells by regulating various cellular processes (1). Microbial metabolites and other secreted factors affect miRNA/gene expression profiles in cells present in the tumor microenvironment. In turn, tumor cells affect the microbiota composition of the stromal and tumor-infiltrating immune cells through the shedding of epithelial cells and/or secreting extracellular vesicles (EVs) containing miRNAs (2). The tumor-miRNAs alter the microbiota composition by affecting the gene expression of the microbiota and by delivering cancer-secreted metabolites (3). The tumor-derived miRNAs also have a role in regulating stromal and tumor-infiltrating immune cells by affecting gene expression through miRNAs delivered in EVs (4). Such interactions will finally create a favorable microenvironment for tumor cells that include angiogenesis, immune evasion, and microbiota composition (5).

Bacteroides fragilis and *Fusobacterium nucleatum* has been found. *F. nucleatum* is commonly found in the human oral microbiota and is frequently associated with gum diseases; it is, however, not commonly present in the gut microbiota. Furthermore, high levels of anti-*Fusobacterium* IgA and IgG have been detected in the sera of CRC patients opening the potential for diagnostic biomarkers[27]. A higher abundance of the Bacteroidetes phylum and a lower abundance of the Firmicutes phylum have been observed in CRC patients. A recent meta-analysis of various CRC microbiota datasets found, for the tissue-associated microbiota, a consistently higher abundances of *F. nucleatum*, *Parvimonas*, and *Streptococcus*; 9 studies in that meta-analysis found a consistently lower abundance of *Faecalibacterium* and *Ruminococcaceae*[6].

However, because the microbiota works as a community and can change in composition in response to many environmental factors, it is difficult to properly model the human microbiota in mice.

1.2.1 Microbial Metabolites

In a normal colon, the microbiota produces a vast number of metabolites. Some of them, including vitamin K, biotin, and short-chain fatty acids (SCFAs), is essential for maintaining homeostasis in the colon microenvironment[28]. In fact, the major energy source (~70%) required by colon epithelium is butyrate, which is produced by the microbiota through fermentation of complex carbohydrates.

Without the microbiota, the colon epithelium undergoes autophagy and fails to maintain its normal structure and function[29]. Similarly, mice lacking a microbiota (i.e., germ-free mice or those treated by broad-spectrum antibiotics) develop significantly fewer tumors in the colon[24,25,30]. However, in humans, using broad-spectrum antibiotics to treat CRC is not feasible, because of the risk of introducing harmful and highly resistant secondary infections such as *Clostridioides difficile*.

In our current understanding, a few main classes of bacterial metabolites play a key role in the pathogenesis of CRC and the immune microenvironment. These metabolites include SCFAs, polyamines, secondary bile acids, and phytochemicals. Their role in CRC has been extensively reviewed and documented[28,31,32,33].

1.2.2 Microbial Factors

In the past few years, researchers have found specific proteins in the gut microbiota, such as FadA and Fap2 proteins from *F. nucleatum* and *B. fragilis* toxins, that play a role in CRC pathogenesis[5,13,16,19,21,26,34].

Using the *Fap2* virulence factor, *F. nucleatum* uniquely binds with the D-galactose-β(1-3)-N-acetyl-D-galactosamine (Gal-GalNAc) carbohydrate moiety expressed on the tumor surface of CRCs[19]. Once it localizes to the CRC microenvironment, it targets the *Wnt/β-catenin* signaling pathway by binding, via

association with the FadA virulence factor, to the E-cadherin protein on the cell surface[21]. The *Wnt/β-catenin* signaling pathway is critical during tumor initiation, tumor migration, and metabolic reprogramming[35–37,38]. The *B. fragilis* toxin (bft) also targets the same *Wnt/β-catenin* signaling pathway[39]. The bft virulence factor is able to bind to the E-cadherin protein, similar to that of FadA, but additionally cleaves the protein, which can alter the intestinal tight-junction function[40]. The *Wnt/β-catenin* pathway is a major signaling pathway that controls the expression of many important tumor-suppressor-related genes, including *MYC*. The transcription factor *MYC*, transactivates miRNAs, such as the miR-17-92 cluster, that are highly expressed in CRC[41–44].

Additionally, *F. nucleatum* can also induce CRC cell proliferation by upregulating miR-21, via activation of the nuclear factor kappa-light-chain-enhancer of activated B cells (*NF-κB*) pathway via toll-like receptor 4 (*TLR4*) signaling[45]. *Escherichia coli* harboring the *pks* genomic island also plays an important role in CRC. When CRC cells come in contact with the colibactin genotoxin produced by *E. coli*, the cells undergo cellular senescence[46,47]. This process is mediated by the cellular upregulation of miR-20a-5p, which results in the downregulation of sentrin-specific protease 1 (*SENP1*). This process then alters p53 small ubiquitin-like modifier (SUMO)ylation, which has been shown to affect the growth and metastasis of tumor cells[48]. It has also been found that biofilms can promote CRC with increases in *IFN-γ* and *IL-17* as well as inhibition of *NF-κB* which have been shown to have immunosuppressive effects in pre-clinical models through reduced expression of anti-inflammatory cytokines[49–51].

In addition to virulent factors, many bacteria also produce beneficial factors that can reduce inflammation and modulate the immune system. In germ-free mice, early studies found impaired intestinal immune systems, which were amenable to treatment[52]. Specifically, the *B. fragilis* polysaccharide A (PSA) is one such immunomodulatory factor that maintains the proper function of CD4+ T cells[53]. Several other polysaccharides produced by *B. fragilis* are also beneficial in maintaining proper immune function. Immunization with *B. fragilis* polysaccharides, or the adoptive transfer of T cells specific to *B. fragilis*, can even boost the treatment effect of anti-cytotoxic T-lymphocyte antigen 4 (CTLA-4) immunotherapy[54]. The seemingly conflicting role of *B. fragilis* within gut bacteria is only the tip of the iceberg in current microbiota research and the fine and highly complex balance between functions.

1.3 MicroRNAs and Colorectal Cancer

MicroRNAs are small non-coding RNAs (about 22 nt) that play an important role in regulating and fine-tuning gene expression[55]. In mammalian cells, miRNAs regulate gene expression through posttranscriptional modifications in two distinct, albeit paired, mechanisms. First, if the miRNA has an extensive complementary binding site in the messenger RNA (mRNA) target, then it will guide the RNA-induced silencing complex (RISC) to cleave the mRNA, thus inhibiting translation. Second, if the miRNA only partially binds to the 3’

untranslated region (3'UTR) of the mRNA, then the miRNA-RISC will act to repress mRNA translation[56]. Both mechanisms lead to the decreased translation of mRNAs, which alters their respective downstream functions. Because miRNAs can act upon mRNA targets with limited complementarity, each miRNA can target a wide range of mRNAs in mammalian cells and each mRNA can be targeted by numerous miRNAs. More than 30% of human genes are estimated to have conserved binding sites in the 3'UTR[57]. Clearly, given this vast and enormously complex regulatory network, miRNAs are immensely important in regulating critical cellular processes. We are only now beginning to understand the sophisticated crosstalk of miRNAs, not only with each other but with the myriad of target mRNAs.

Previous studies have identified numerous aberrant miRNA expression patterns in CRC[58–62]. Specifically, the miR-17-92 cluster, miR-21, miR-182, and miR-503 are consistently overexpressed in tumor (vs. normal) tissues [3,41,42,59,63–72]. Any alteration(s) in expression levels of these miRNAs could, in turn, affect a wide array of downstream gene targets. Together, these miRNAs regulate all aspects of tumor pathobiology, including (i) altering tumor metabolism; (ii) promoting cell proliferation; (iii) stimulating angiogenesis; (iv) down-regulating tumor-suppressor genes; (v) promoting evasion of immune surveillance, and (vi) creating a favorable tumor microenvironment that promotes invasion and metastasis.

Our laboratory previously reported that, during the adenoma-to-adenocarcinoma transition, miR-182 and miR-503 were sequentially

overexpressed and targeted the tumor-suppressor *FBXW7* gene[70]. Other researchers have observed, during CRC transformation, an increased expression of the miR-17-92 cluster and miR-21[41,73]. In CRC adenocarcinoma, members of the miR-17-92 cluster target transforming growth factor-beta (*TGF-β*), which in turn stimulates angiogenesis in the tumor microenvironment, thus promoting tumor growth [71]. Additionally, miR-19, a member of the miR-17-92 cluster, downregulates the expression of the tumor-suppressor phosphatase and tensin homolog (*PTEN*), thereby activating the protein kinase B (*AKT*)/mammalian target of rapamycin (*mTOR*) pathway in tumor cells[74]. The *AKT*/*mTOR* pathway is the main metabolic sensing pathway, responsible for regulating glucose transport into cells[75]. Since glucose is the main fuel source of CRC cells and activated *AKT*/*mTOR* pathway promotes tumor cell proliferation[76].

The tumor-suppressor *PDCD4* gene, which is commonly downregulated in CRC, is a target of miR-21[68]. Inhibiting the *PDCD4* gene can lead to an increase in the metastasis potential of tumor cells. Another important pathway commonly altered in CRC tumors is the *Wnt/β-catenin* pathway[58,63]. Dozens of miRNAs have been shown to extensively regulate the genes involved in the *Wnt/β-catenin* pathway[58]. The *Wnt/β-catenin* signaling pathway is critical during tumor initiation, tumor migration, and metabolic reprogramming[35–38]. The role of the *Wnt/β-catenin* signaling pathway in CRC has been previously reviewed[38].

The complex microenvironment of the CRC tumor also involves stromal cell and immune cell fractions, which can be regulated by cancer-derived miRNAs[77–79]. Studies have found that the miR-17-92 cluster, commonly

overexpressed in CRC cells, is also upregulated in CRC stromal cells[69,73,80]. Strikingly, these miRNAs are not only endogenously produced by stromal cells but also packaged in the microvesicles of tumor cells, and then delivered to stromal cells[81,82]. Similar intracellular regulation mediated by miRNAs is also found in immune cell fractions[83]. Additionally, endogenous miRNA dysregulation is prevalent in CRC immune cell fractions, usually as a downstream effect of tumor-secreted factors such as cytokines and chemokines[84–86]. Collectively, this evidence suggests that miRNAs are important in regulating tumor cells, in addition to maintaining the tumor microenvironment. It is clear that the relationship between miRNAs and CRC is multifaceted, interrelated, and highly complex.

1.4 Tumor nutrient microenvironment changes

The development of CRC entails a complex interplay between the epithelial cells, the microbiota, and the immune system in the tumor microenvironment[34], and multiple signaling pathways play critical roles in both tumorigenesis and tumor progression. Tumor metabolism changes have been well studied and characterized. One of the hallmarks is an increase in glycolysis as the primary energy source, known as the Warburg effect[87]. Several studies have found altered metabolite levels in both tissues and stools of CRC patients. In tissue samples (as compared with adjacent normal tissues), glucose levels were significantly lower, whereas levels of lactate and fatty acids were significantly higher[88–90]. In stool samples of CRC patients, amino acid levels were

higher than normal; levels of fatty acids were lower[88–90]. These nutrient composition changes in CRC patients correspond to the tumor's increased need for glucose for energy and fatty acids for proliferation.

miRNAs play a critical role in regulating the metabolism of CRC patients and in sustaining the needs of tumor cells **(Figure 1.2)**. The extracellular glucose is first transported into cells via the glucose transporter 1 (*GLUT1*) receptor, which is a downstream target of the mammalian target of the rapamycin (*mTOR*) gene. In CRC patients, the *mTOR* gene is regulated by miR-144. Higher expression of miR-144 inhibits expression of the mTOR gene, leading to reduced glucose uptake by the tumor cells; thus, higher expression of miR-144 is associated with a good prognosis for CRC patients[91].

After the glucose is transported into the cytosol, it undergoes glycolysis—a process regulated by the alternative splicing of pyruvate kinase (*PK*). Higher levels of 2 PK isoforms, M1 (*PKM1*) and M2 (*PKM2*), in cells, will lead to increased glycolysis, instead of oxidative phosphorylation[92–94]. These studies have found that overexpressing miR-124 (another regulator of the mTOR gene) in CRC cells can lead to higher *PKM1*:*PKM2* ratios, thus inhibiting glycolysis and controlling tumor cell growth. The end product of glycolysis, pyruvate, will then be metabolized into lactate by lactate dehydrogenase A (*LDHA*), which is commonly upregulated in CRC patients. *LDHA* is a rate-limiting enzyme of glycolysis, so lost *LDHA* expression is associated with a decrease in adenosine triphosphate (ATP) production and cell proliferation[95]. In CRC cell lines, various miRNAs—including

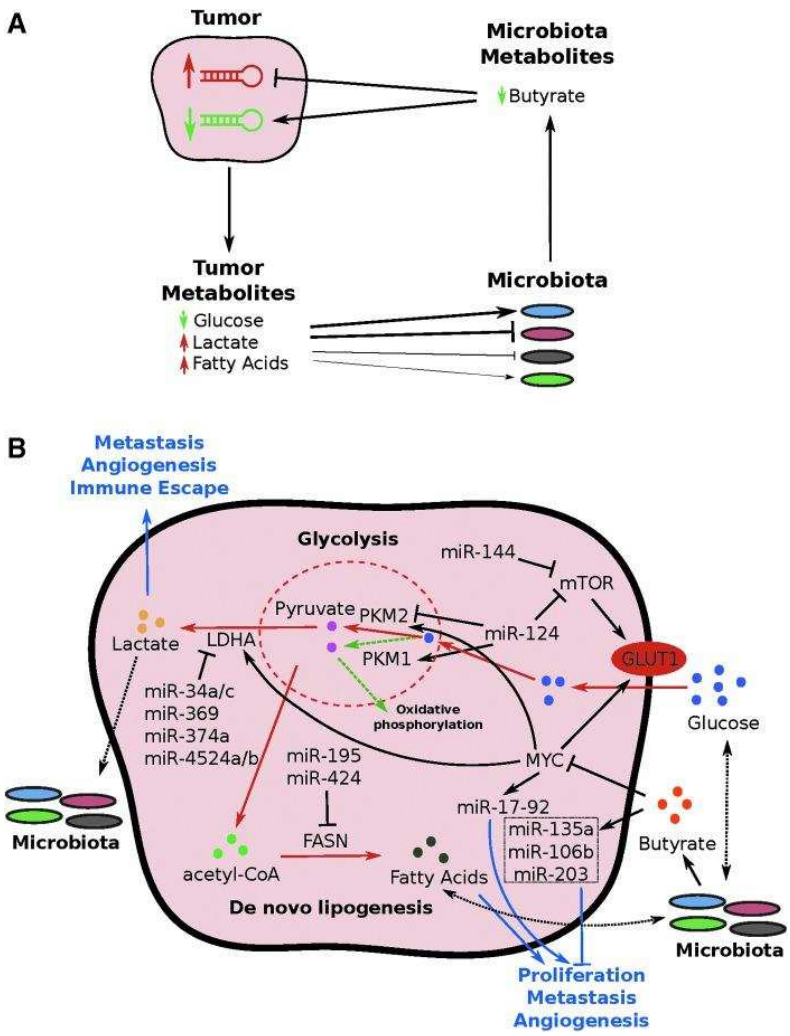

Figure 1.2. Tumor–microbiota metabolic interactions.
(A) An overview of the tumor–microbiota metabolic interactions. The thickness of the line connecting tumor metabolites and microbiota shows the relative effects of tumor metabolites on microbiota composition. (B) A curated map of miRNA-mediated tumor–microbiota metabolic interactions. Red lines indicate the pathway upregulated in colorectal cancer relative to normal tissue. Blue lines indicate the overall downstream effects of the miRNAs or metabolites. Dotted green lines indicate the oxidative phosphorylation common in normal cells. Dotted black lines indicate potential effects. miRNA, microRNA.

miR-34a/c, miR-369-3p, miR-374a, and miR-4524a/b—have been shown to inhibit *LDHA* expression[95]. The lactate produced by the tumor cells can function as signaling molecules that further affect tumor cell metastasis, angiogenesis, and immune escape[96].

In addition to altered glucose metabolism, CRC cells also have altered macromolecule metabolism. We postulate the reason for the higher levels of fatty acids in CRC patients' tissue samples (relative to their stool samples) is the increased need for membrane synthesis to support cell proliferation[88–90]. One of the most important genes controlling this pathway is the fatty-acid synthase (*FASN*) gene. The enzyme encoded by the *FASN* gene is critical for controlling the synthesis of lipids, a process required for cell membrane formation. In breast cancer and osteosarcoma, studies have found that miR-195 and miR-424 target the *FASN* gene, thus inhibiting cell proliferation, invasion, and metastasis[97–99]. Both of those miRNAs are significantly upregulated in CRC tissues, according to the Cancer Genome Atlas (TCGA) dataset, suggesting such microRNA-mediated lipogenesis may also happen in CRC. Additionally, because *FASN* is potentially important to T-cell immunity[100], the dynamics of the miR-424/*FASN* axis in tumor and immune cell function are currently being worked out. Other pathways with downstream effects on metabolism, such as *PTEN*, and *AKT*/*PI3K* are also modulated by miRNAs[65,101–104]. Based on current evidence, it is clear that aberrant miRNA expression in CRC cells profoundly alters the nutrient composition of the tumor microenvironment.

1.4 Host Regulation of Microbiota Mediated by MicroRNAs

In reestablishing germ-free mice with normal microbiota, studies have found altered intestinal miRNA profiles, suggesting that the microbiota regulates host miRNA expression[105,106]. Moreover, the responses of intestinal cells to facilitating the microbiota process depend on the cell type, and intestinal epithelial stem cells are especially sensitive to microbiota reestablishment[106]. Because miRNAs are highly stable, several studies in the clinical arena were able to detect higher levels of miR-21 and miR-92a, among other miRNAs, in the fecal samples of patients with CRC[66,107,108]. This finding facilitated the development of a noninvasive CRC screening method and delineated the potential role of miRNAs in interacting with the trillions of microbes in the human gut.

Intestinal miRNAs develop from two main sources, including the host and the food[109,110]. The intestinal epithelial cells are the main contributors of host-derived miRNAs, either via shedding of cells or excretion of exosomes. Evidence has shown that miRNAs from food can be absorbed by the host and can affect host gene expression[111–113]. But certain food-sourced miRNAs remain stable in the digestive tract and reach the intestines[114,115]. This evidence suggests that miRNAs can mediate cross-species regulation. The idea remains nascent, so insight into how miRNAs mediate host-microbiota interactions is still limited. Liu et al. first demonstrated such regulation, showing that miRNAs present in the feces can regulate gene expression and growth of bacteria[109]. Specifically, they

found that mice lacking the Dicer gene, which enables mature miRNA processing, had different microbiota profiles than wild-type mice. More importantly, the study reported that hsa-miR-515-5p promoted the growth of *F. nucleatum* in vitro by targeting the 16S ribosomal RNA (rRNA) gene.

Notably, however, hsa-miR-515-5p shows very low expression levels in CRC tumors, so they are not significantly different from normal tissue. Thus, interactions between hsa-miR-515-5p and *F. nucleatum* might not be significant in CRC pathogenesis. However, more importantly, this study found that fecal miRNA transplantation restores fecal microbiota composition in mice with Dicer gene knockout. Several recent studies found that fecal microbiota transplantation (FMT) offers a potential therapeutic benefit that enables an immunotherapeutic response[30,116–119]. Based on growing evidence, it is plausible that fecal miRNAs play an important role in modulating the CRC microbiota as well as immunotherapy responses.

Recently, Teng et al. demonstrated that miRNAs encapsulated in plant-derived exosome-like nanoparticles (ELNs) can enter bacteria and alter bacterial genes[110]. The process for bacterial uptake of ELNs is determined primarily by the lipid composition of the outer membrane. They found that ELNs enriched with phosphatidylcholine were preferentially taken up by the *Ruminococcus* spp., whereas ELNs enriched with phosphatidic acid (PA) were primarily taken up by *Lactobacillus rhamnosus*. After the ELNs are taken up by specific bacteria, the miRNA contents are released into bacterial cells. Teng et al. also found that mdo-miR7267-3p encapsulated in the PA-enriched ELNs targets the *Lactobacillus*

monooxygenase *ycnE*, which then increases its production of indole-3-carboxaldehyde (I3A). The I3A metabolite then promotes interleukin-22 (*IL-22*) production and helps repair damaged colon mucosa[120].

There is developing evidence to support the notion that host or exogenous miRNAs might be biologically active in bacteria, thereby affecting bacterial gene expression. Although small RNAs similar to miRNAs exist in bacteria and function similarly to miRNAs, it remains unknown how miRNAs function in bacteria[121]. Several studies have reported that exogenous miRNAs from plant or animal sources can be taken up by human cells and exert biological functions[111–115,122–124]. Additional studies are required to ascertain whether or not miRNAs can indeed affect bacteria and to delineate the precise mechanism(s).

1.5 Conclusions and Perspectives

With thousands of bacterial species living in the human digestive tract, it is becoming quite evident that they profoundly affect human health. Our review of the recent literature regarding CRC underscores a complex metabolic interplay between the host and its microbiota, mediated in part by miRNAs. Based on the current literature, we offer five major points in host-microbiota interactions in CRC **(Figure 1.1)**:

1. The CRC microbiota has reduced representation of beneficial bacteria. These beneficial bacteria produce metabolites and other factors that can

potentially slow CRC progression, in part via the modulation of miRNAs that regulate tumor and immune cells.

2. Dysregulation of miRNAs in tumor cells can affect the survival, or the gene expression, of certain members of the microbiota.
3. Dysregulated miRNAs in tumor cells can be packaged and delivered to both stromal and immune cell fractions, creating a more favorable microenvironment for tumor cells.
4. Overrepresentation of oncogenic bacteria among the CRC microbiota can modulate tumor and immune cells, as well as the broader tumor microenvironment, thereby resulting in a more favorable condition for tumor growth.
5. This negative feedback loop perpetuates CRC progression.

Based on both experimental and computational data, we believe that miRNA and metabolites mediate and critically influence host-microbiota interactions. Clearly, they are a major part of a complex web of highly dynamic interactions. While other factors in the CRC microenvironment could also play important roles, in the current dissertation, we focus our investigation on the role of miRNA and metabolites in mediating the interactions between host and microbiota.

1.9 PUBLICATIONS

This chapter has been modified (with permission) from the published articles[26,125,126]:

26. Yuan, C., Steer, C. J. & Subramanian, S. Host−microRNA−microbiota interactions in colorectal cancer. *Genes* (Basel) 10, (2019). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6523287/

125. Yuan, C. & Subramanian, S. MicroRNA Mediated Tumor-Microbiota Metabolic Interactions in Colorectal Cancer. *DNA Cell Biol* (2019).

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6477581/

126. Yuan, C. et al. Tumor models to assess immune response and tumor-microbiome interactions in colorectal cancer. *Pharmacol. Ther.* 231, 107981 (2022).

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8844062/

CHAPTER 2: MUCOSAL MICROBIOTA AND METABOLOME ALONG THE INTESTINAL TRACT REVEAL A LOCATION-SPECIFIC RELATIONSHIP

2.1 INTRODUCTION

The human intestinal tract harbors trillions of microorganisms, termed the microbiota, which includes thousands of bacterial species[127]. It has become evident that the gut microbiota is important in regulating and maintaining the health of the host and is implicated in many diseases, such as obesity and several cancers[3,23,128–130]. Despite numerous studies indicating the important roles of microbiota in diseases, most studies have primarily focused on variations in the taxonomic composition of the microbiota. The underlying metabolic features associated with the host-microbiota interaction, however, remain unclear for most diseases.

The gut microbiota produces a vast amount of metabolites. Some metabolites, such as vitamin B, vitamin K, bile acids, and short-chain fatty acids (SCFAs), are essential to maintaining homeostasis in the colon[28,131,132]. The most direct and active metabolic interaction between the host and its microbiota is in the large intestine, and the vast majority (~70%) of energy required by the normal colon epithelium comes from butyrate produced by the microbiota through fermentation of complex carbohydrates[133]. Without a functional microbiota, the colon epithelia undergo autophagy and fail to maintain normal structure and

function[29]. Moreover, the metabolic interactions between the host and its microbiota have widespread implications throughout the body[131]. For example, the obesity-associated microbiota has been shown to possess the increased metabolic capability to harvest energy from food[134,135], and the metabolism of L-carnitine by the gut microbiota has been shown to promote atherosclerosis[136]. These studies suggest potential metabolic shifts of the microbiota, either in response to or responsible for the host metabolic state[135].

The mucosal host-microbiota metabolic interactions along a healthy human intestinal tract are largely unknown. Although the microbiota and metabolome variations along the intestinal tract have been investigated in rodents and other animals, the dietary and anatomical differences between humans and these animals render these data less informative for humans[137–142]. Here, we investigated the microbiota and metabolome profiles along the intestinal tracts of healthy baboons (*Papio anubis*), a family of Old World monkeys. We collected tissue samples from the duodenum, jejunum, ileum, cecum, proximal colon, and distal colon. Amplicon sequencing of the 16S rRNA gene (16S-Seq) was used to identify the mucosal surface microbiota composition. We also performed untargeted metabolomics on the immediately adjacent tissues to profile the tissue metabolite contents.

2.2 **MATERIALS AND METHODS**

The tissue samples were collected via tissue sharing postmortem, which is exempt from Institutional Animal Care and Use Committee (IACUC) review. The cohort included 10 adult purpose-bred female olive baboons (*Papio anubis*) modeling anterior cruciate ligament (ACL) injury and subsequent repair using regenerative medicine techniques. The animals were between 6.5 and 15.6 years old (median, 9.3 years) and weighed between 14.4 and 24.9 kg (median, 20.1 kg). They were housed in pairs or housed in protected contact with compatible conspecifics. Baboons had free access to water and were fed identical diets that included biscuits (Harlan primate diet 2055C; Harlan Teklad) based on body weight and daily enrichment with fresh fruits, vegetables, grains, beans, nuts, and a multivitamin preparation. Semiannual veterinary physical examinations were performed on all animals. Animals participated in an environmental enrichment program designed to encourage sensory engagement, enhance foraging behavior and novelty seeking, promote mental stimulation, increase exploration and play and activity levels, and strengthen social behaviors, providing opportunities for animals to increase time spent on species-

typical behaviors. Baboons were trained to cooperate with medical procedures, including hand feeding and drinking, shifting into transport cages for sedation, and targeting or presentation for examination. Animals were euthanized via barbiturate overdose (Beuthanasia-D ≥86 mg/kg of body weight intravenously), and tissue procurement was performed postmortem. No oral medications were used for at least 6 months prior to tissue collection. Tissue sections (approximately 1 cm by 1 cm) from six different sites that included the duodenum, jejunum, ileum, cecum, proximal colon, and distal colon were collected from each animal, a total of 60 samples, using clean technique, snap-frozen in liquid nitrogen, and then stored at –80°C.

2.2.1 16S-Seq and sequence analysis.

Total DNA was extracted from approximately 250 mg of tissue using DNeasy PowerSoil kit (catalog no. 12888; Qiagen, Valencia, CA) following the standard protocol. Sequencing libraries were created by the Mayo Clinic Genome Analysis Core (Rochester, MN). Briefly, the V3-V5 region of the 16S rRNA gene was amplified with multiplexing barcodes using PCR (V3-341F,

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGAGGCAGCAG; V5-926R, GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCGTCAATTCMTTTRAGT). The libraries were then pooled and size selected between 700 and 730 bp using a LabChip XT (PerkinElmer, Waltham, MA). Sequencing was performed on a single lane of a MiSeq sequencer (Illumina) using paired-end mode. On average, 64,937 quality reads (between 9,901 and 118,288) were generated per library. The sequencing results were analyzed using the IM-TORNADO2 pipeline [143]. Alpha- and beta-diversity metrics were analyzed using QIIME v1.9.1[144]. The unfiltered OTU table is available in Appendix (**Appendix Table AT1**) in the supplemental material. Linear discriminant analysis of effect size (LEfSe) was used to determine differences in the relative abundances of taxa among tissue sites[145].

The beta-diversity between tissue locations was analyzed by performing principal-coordinate analysis (PCoA) using both weighted and unweighted UniFrac distance metrics. The unweighted UniFrac distance considers only the presence and absence of a certain OTU, while the weighted UniFrac distance will consider the abundance; thus, these metrics can give an overview of the microbial structure differences of different tissue locations[146–148].

2.2.2 Metabolite extraction.

Metabolites were extracted from the immediately adjacent tissue that was used to generate 16S-Seq. There was an insufficient amount of duodenum tissue from animal B09 to perform untargeted metabolomics, so it was not analyzed. Approximately 15 mg of tissue was used to extract metabolites. The tissues were first ground into fine powder using CryoGrinder (OPS Diagnostics) on dry ice. The tissues were then suspended in 20 µl of 80% methanol per 1 mg of tissue weight. The mixture was then homogenized using a probe sonicator at 10% amplitude for 15 s, with 1-min rest on ice after 5 s of sonication. The sonicated samples were then centrifuged at 14,000 × *g* for 10 min at 4°C. The supernatant from the centrifugation contained the metabolites and was saved at –80°C before drying. The tissue pellets were then further processed for additional metabolite extraction. They were first suspended in 10 µl of 80% methanol per 1 mg of original tissue weight and sent through high-pressure cycling on a Barocycler NEP2320 (Pressure Biosciences). The high-pressure cycling protocol includes 60 cycles of 20 s of 35,000 lb/in^2 pressure, followed by 10 s of 0 lb/in^2 at 4°C. After pressure cycling, the samples were again centrifuged at 14,000 × *g* for 10 min at 4°C, and the supernatants were pooled with the previously extracted metabolites. Finally, the metabolites were dried under a nitrogen stream.

2.2.3 Untargeted metabolomics.

The dried metabolites were first suspended in 15 μl of 0.1% formic acid per 1 mg of the original tissue weight. The suspensions were then separated for analysis using a C_{18} reverse-phase column and hydrophilic interaction liquid chromatography (HILIC) column. The reverse-phase analysis results in separation of larger nonpolar molecules such as steroid-like compounds, certain amino acids, phospholipids, and other lipids, while the HILIC analysis separates hydrophilic compounds such as amino acids and member of the citric acid cycle and glycolysis pathways. The samples were analyzed using reverse-phase positive mode (nonpolar interaction) separation and HILIC analysis (polar interaction) separation before analysis with Q Exactive LC-MS/MS quadrupole Orbitrap (Thermo Scientific). The reverse-phase analysis was performed in positive mode ionization with an additional proton (+1.0073) added. For HILIC analysis, the negative ionization mode was used with one additional proton (−1.0073) removed. Since salts are present, compounds may occasionally form as a sodium salt (neutral mass plus 21.9944) for the positive mode or as a chloride salt (neutral mass plus 34.9688) for the negative mode. Samples were

loaded and analyzed in random order, and quality control samples were analyzed at regular intervals to eliminate extraneous signals. The untargeted metabolomics were performed by the University of Minnesota Center for Mass Spectrometry and Proteomics.

2.2.4 Metabolomic data analysis.

The data were processed using Progenesis QI software (Thermo). The software first aligns all the features obtained in all the runs and then assigns intensity measures for features found in all the runs. The raw data were further processed by filtering for fidelity of individual feature detection using the quality control samples. Only features with a coefficient of variation (CV) of less than 10% overall quality control samples were accepted. Features showing high intensity in background samples relative to the quality control samples and features not present in at least 67% of all samples were removed from analysis per the U.S. Food and Drug Administration recommendation. Each feature is uniquely identified with the mass-to-charge ratio (*m/z*) and the elution time from the column. Features were then assigned to metabolites identified by searching the Human Metabolome DataBase (HMDB) and using databases developed by

the University of Minnesota (**Appendix Table AT2**). Pathway analysis was performed using Ingenuity Pathway Analysis (IPA).

2.2.5 Microbiome-metabolome correlation analysis.

All analyses were performed in R v3.4.4 unless otherwise noted. The Spearman's ranked correlation test with false-discovery rate (FDR) adjustment was used to test the microbiome-metabolome correlation[149]. The microbiome OTU data and metabolomic data were first combined and filtered to remove low-abundance OTUs and metabolites (appearing in less than 50% of samples). The Spearman's ranked correlation test was calculated using the *cor. test* function. The *P* values were then adjusted using the *p.adjust* function before filtering for significant correlations. PERMANOVA was performed using *adonis* function with Bray-Curtis distance and 999 permutations. Procrustes analysis was performed using the *procrustes* function of the *vegan* package in R with principal-component analyses of both the microbiome and metabolome using default options[149].

2.3 Results

2.3.1 Microbiota landscape along the nonhuman primate (NHP) intestinal tract.

We first assessed the baboon intestinal-tissue-associated microbiota composition in 10 baboons using the 16S-Seq method. Baboons were between 7 and 16 years old and weighed 14 to 25 kg at the time of sample collection (**Appendix Table AT3**). We found that the small intestinal (duodenum, jejunum, and ileum) microbiota had significantly lower phylogenetic distance ($P < 1 \times 10^{-5}$, two-tailed *t-test*; **Fig. 2.1A**), Shannon index ($P < 1 \times 10^{-5}$; **Fig. 2.1B**), Chao1 index ($P < 1 \times 10^{-5}$; **Fig. 2.1C**), and observed OTUs ($P < 1 \times 10^{-5}$; **Fig. 2.1D**) compared to the microbiota in the large intestine (cecum, proximal colon, and distal colon)[137, 142]. The patterns of beta-diversity also differed between the upper and lower intestinal sites (**Fig. 2.2A** and **B**; **Fig. 2.3**). In the small intestine, differences in composition did not reflect different tissue sites (permutational multivariate analysis of variance [PERMANOVA] $R^2 = 0.02$, $P = 1$), but significant compositional differences were observed between individual animals ($R^2 = 0.84$, $P < 0.0001$; **Fig. 2.2A**). Conversely, in the large intestine, tissue-specific differences were observed ($R^2 = 0.16$, $P < 0.0005$; **Fig. 2.2B**). However, compositional differences were more strongly driven by the individual host ($R^2 = 0.55$, $P < 0.0001$). This suggests that both host and tissue locations can impact the mucosa-microbiota structure in the intestine[3].

The baboon intestinal tissue-associated microbiota was dominated by the bacterial phyla *Firmicutes*, *Bacteroidetes*, *Spirochaetes*, and *Proteobacteria*, independent of the tissue location (**Fig. 2.2C** and **Fig. 2.2D**). At the phylum level,

seven taxa (*Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Fibrobacteres*, *Lentisphaerae*, *Spirochaetes*, and *Verrucomicrobia*) exhibited location-specific enrichment (**Fig. 2.4**; *P-value* cutoff = 0.05, Kruskal-Wallis test with Dunn *post hoc* test). We next analyzed differences in operational taxonomic unit (OTU) composition to discover whether site-specific bacterial community signatures occurred. We performed linear discriminant analysis (LDA) effect size (LEfSe) and identified 21 taxa (at the genus level) that were characteristic of the small and large intestine (**Fig. 2.5**). Of these 21 taxa, 3 taxa (*Brevinema*, *Dehalobacter*, and *Succinivibrio*) were characteristic of the small intestine.

2.3.2 Metabolomic landscape along the baboon intestinal tract.

We then used Q Exactive LC-MS/MS (liquid chromatography coupled to tandem mass spectrometry) quadrupole Orbitrap (Thermo Scientific) to analyze the tissue metabolome composition in tissue samples immediately adjacent to the tissues used for 16S-Seq. A total of 3,395 compounds were present in at least

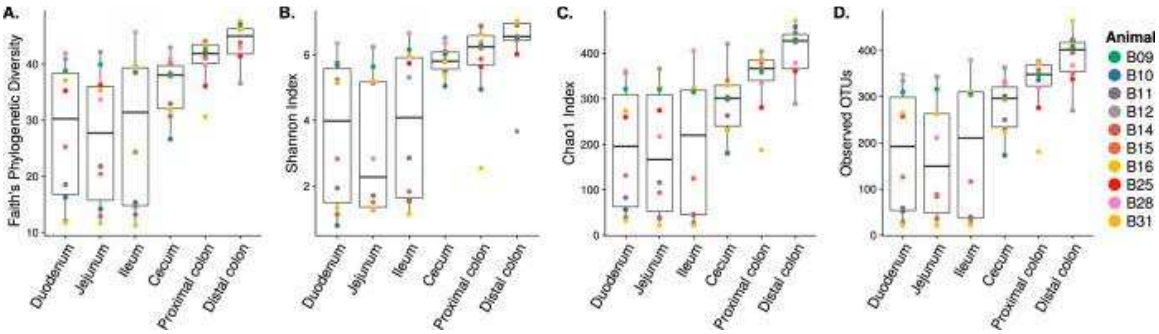

Figure 2.1: Microbiota alpha diversity along the intestinal tract.
(A) Faith's phylogenetic diversity, (B) Shannon index, (C) Chao1 index, and (D) number of observed OTUs.

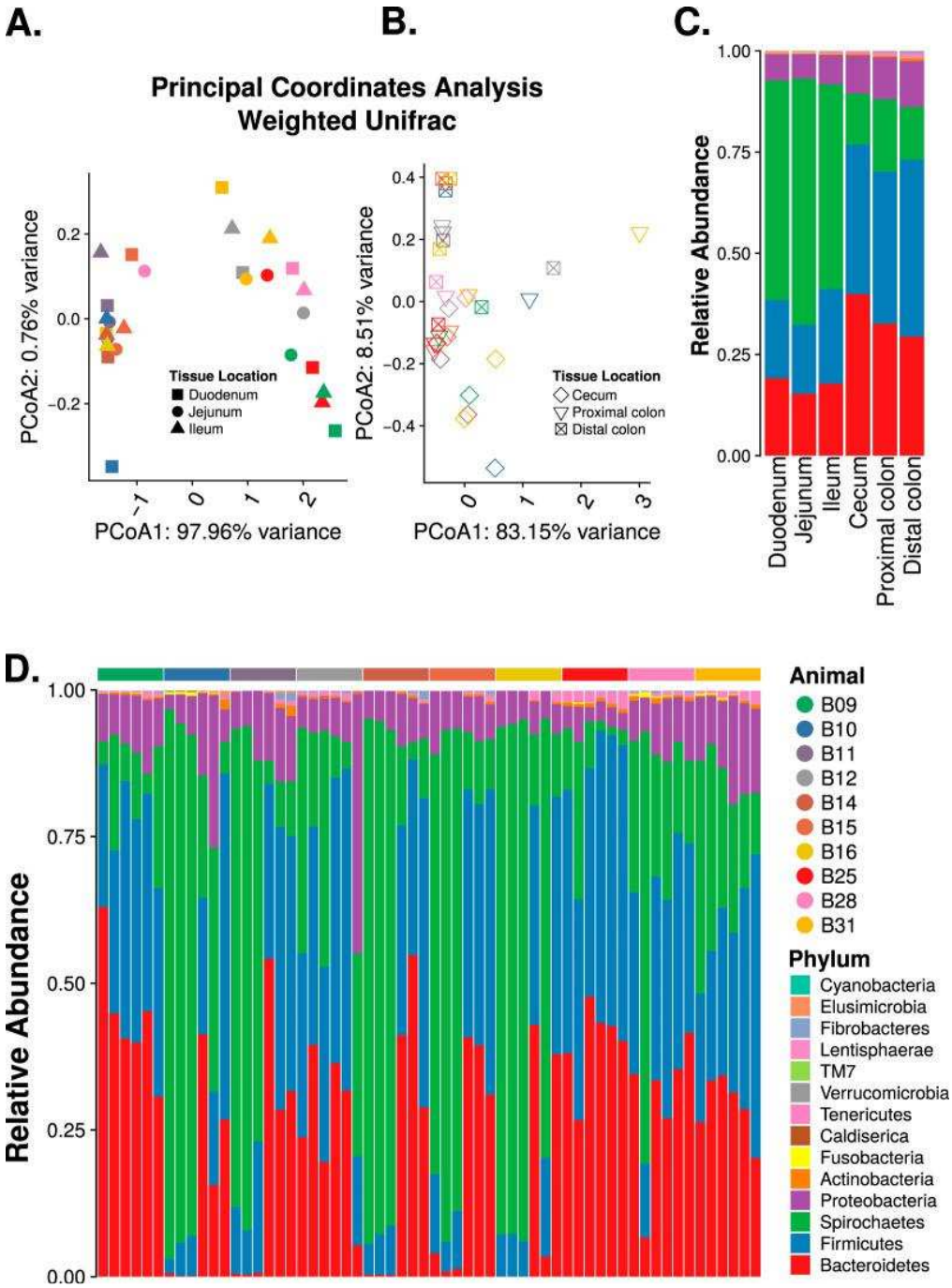

Figure 2.2: Microbiota along the nonhuman primate gastrointestinal tract. Weighted UniFrac principal-coordinate analysis (PCoA) of the upper (A) and lower (B) intestinal samples. Stacked bar plot of bacterial phyla showing the average relative abundance at each of the six tissue locations (C) and at each of the six tissue locations for each sample (D), in the order of duodenum, jejunum, ileum, cecum, proximal colon, and distal colon.

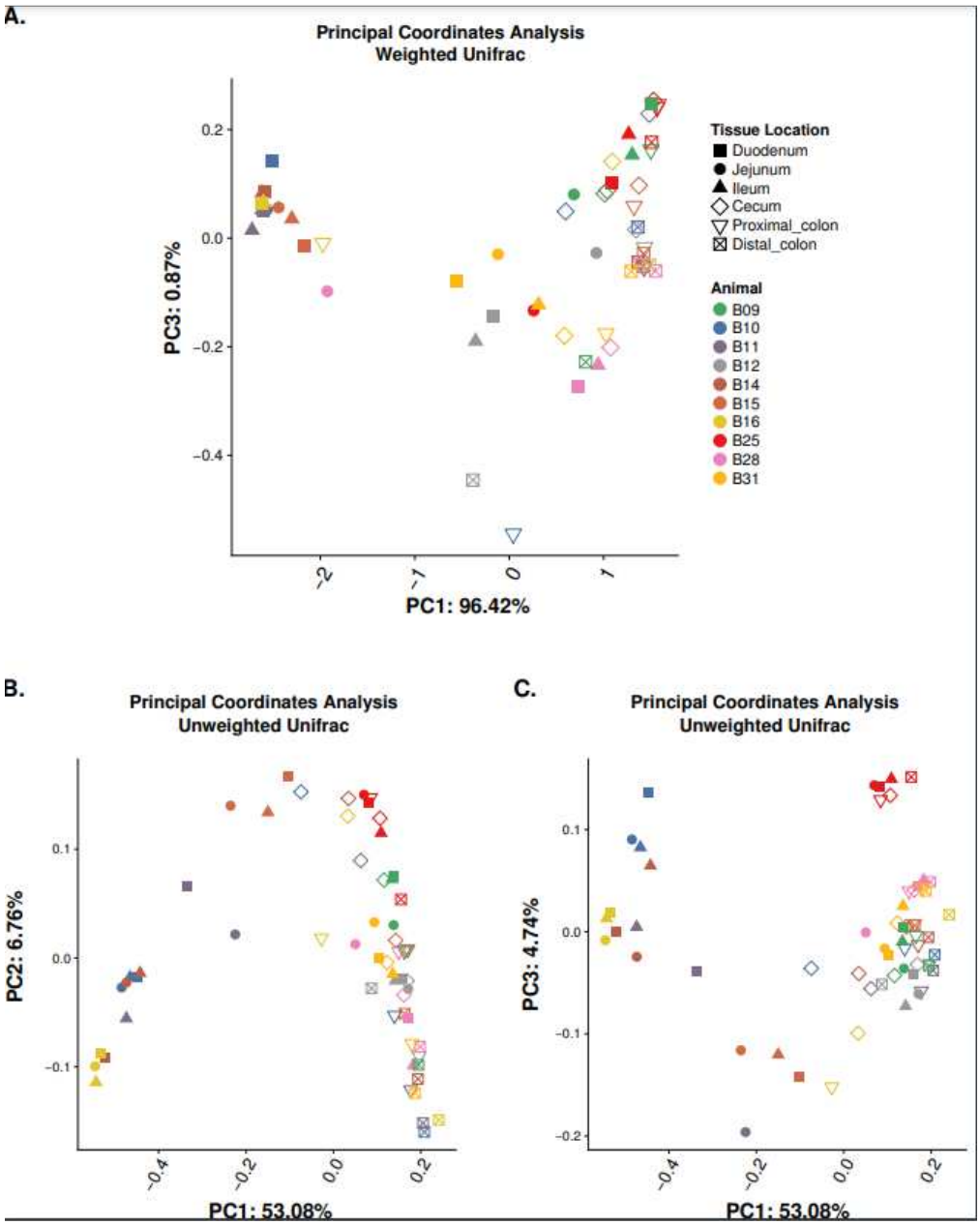

Figure 2.3: Principal-coordinate analysis (PCoA).
(A) Weighted UniFrac PCoA showing PCoA1 versus PCoA3. Unweighted UniFrac PCoA showing PCoA1 versus PCoA2 (B) and PCoA1 versus PCoA3 (C).

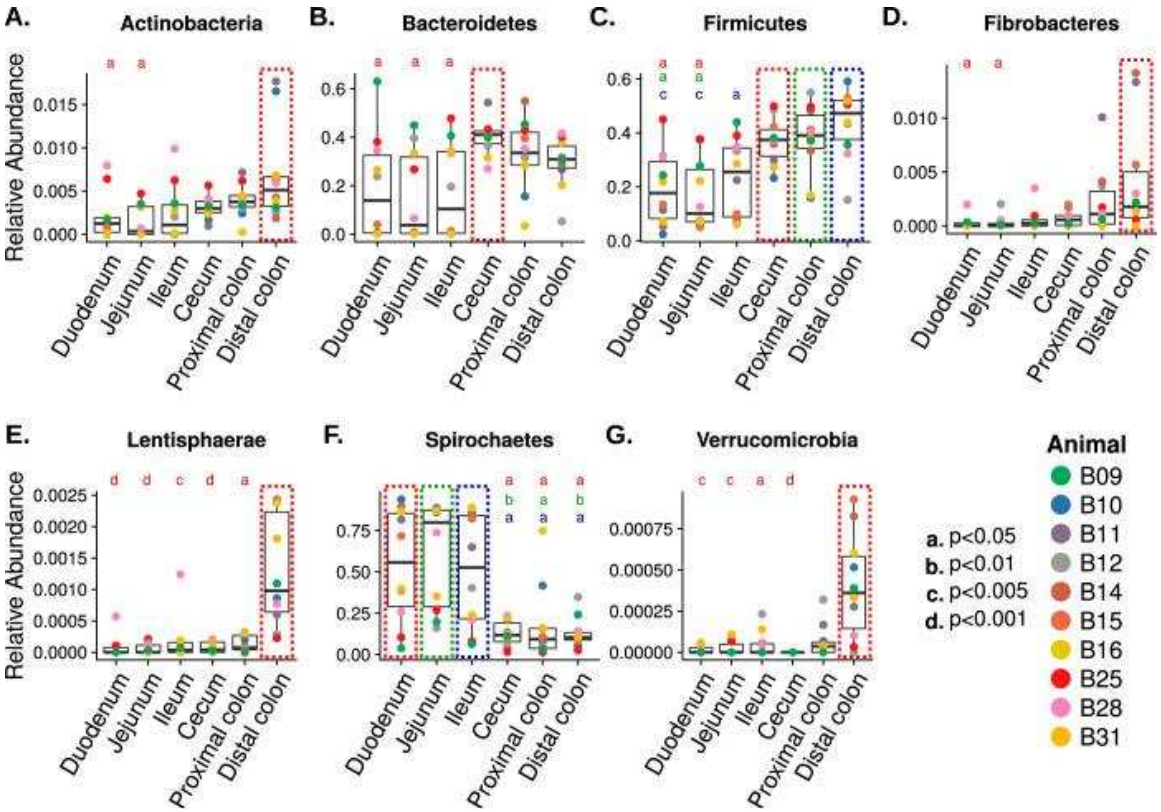

Figure 2.4: Box plot of bacterial phyla with differential abundances across different tissue sites.

Seven bacterial phyla have differential abundance in tissue locations highlighted in the color-dotted box. Statistical significance are indicated by lowercase letters as follows: a, $P < 0.05$; b, $P < 0.01$; c, $P < 0.005$; d, $P < 0.001$.

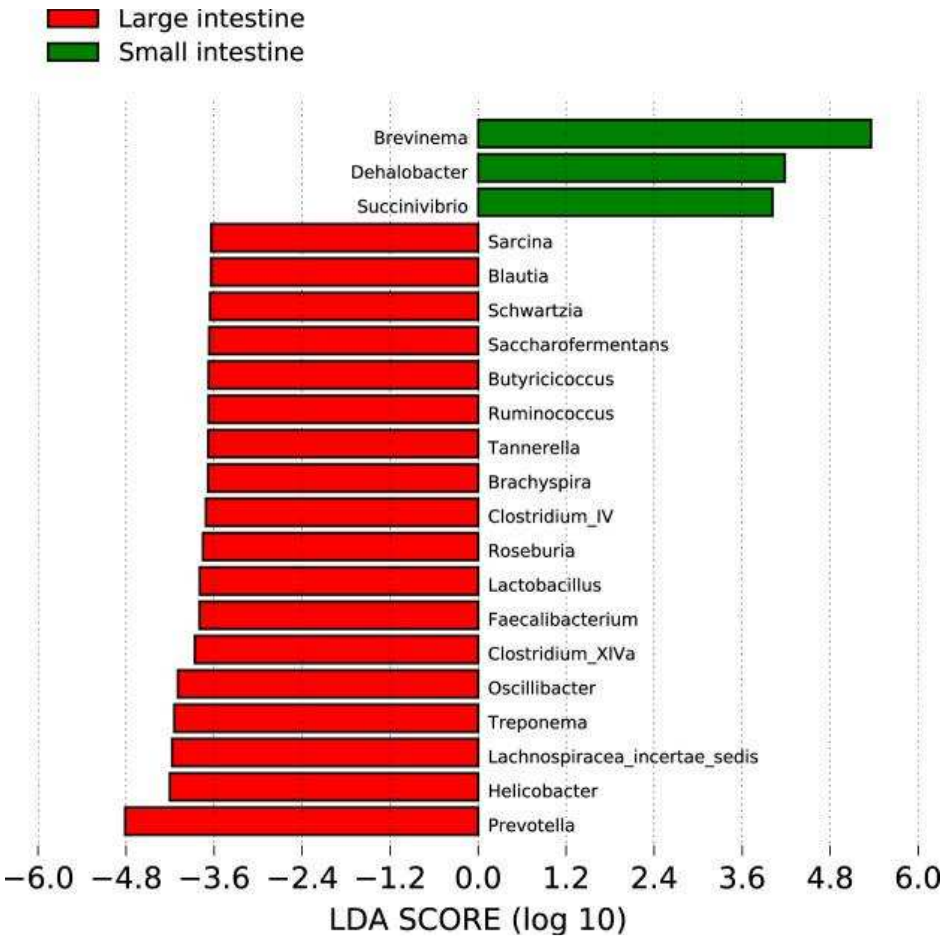

Figure 2.5: Linear discriminant analysis (LDA) effect size (LEfSe).
Thirty-one bacterial taxa have an LDA score (log10) over 3.6. Bacteria with larger effect sizes in the large intestine are shown in red, and bacteria with larger effect sizes in the small intestine are shown in green.

two-thirds of all samples analyzed. After searching against the Human Metabolome Database (HMDB)and in-house libraries generated by the University of Minnesota Center for Mass Spectrometry and Proteomics, a total of 292 compounds were assigned putative identity. We focus on these compounds with assigned identities for further analysis.

We sought first to identify differential metabolites between the small and large intestines. We performed Wilcoxon rank-sum test between metabolites of the small intestine (87) and large intestine (53) and identified 140 compounds with differential abundance. Consistent with previous studies in human and mouse samples, the small intestine contained more amino acids such as aspartic acid, alanine, tyrosine, valine, leucine, and isoleucine, as well as tauro-conjugated bile acids[150,151]. In the large intestine, there was more cholic acid and urobilin, in addition to more-complex metabolites. We then performed pathway analysis using the fold change differences of the differentially abundant compounds between the small and large intestines (**Appendix Table AT3**). Curiously, we found that these compounds are involved in the upregulation of bacterial growth-related pathways (**Fig. 2.6**) in the small intestine. In the large intestine (**Fig. 2.7**), amino acid uptake pathways (**Fig. 2.7B**) and cancer-related pathways (**Fig. 2.7C**) were upregulated.

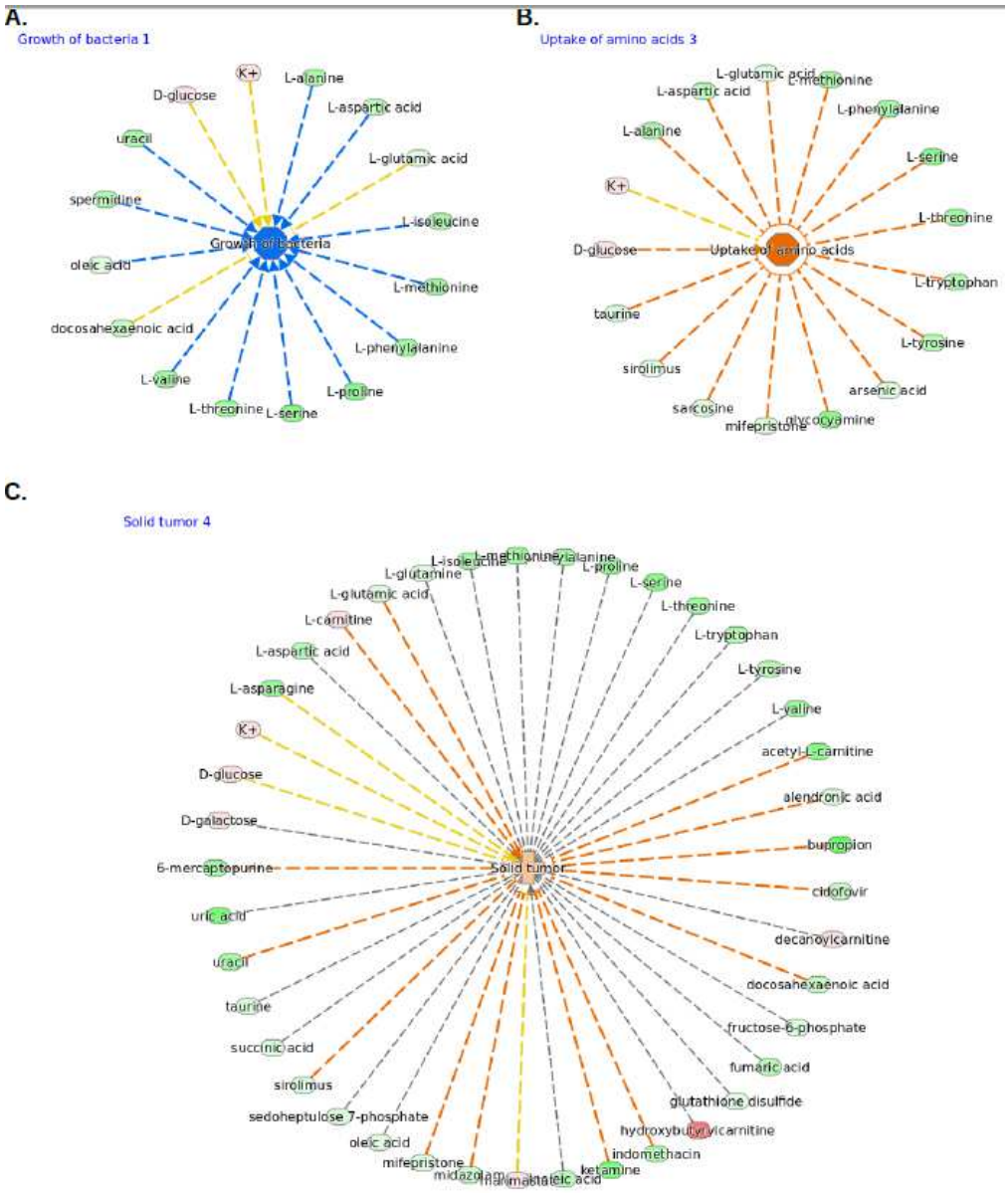

Figure 2.6: Ingenuity Pathway Analysis of differentially abundant metabolites.
(A) Pathways related to growth of bacteria are activated in the small intestines. Pathways related to uptake of amino acids (B) and solid tumor (C) are activated in the large intestines. A blue line indicates activation, an orange line indicates inhibition, a yellow line indicates conflicting evidence, and a gray line indicates association.

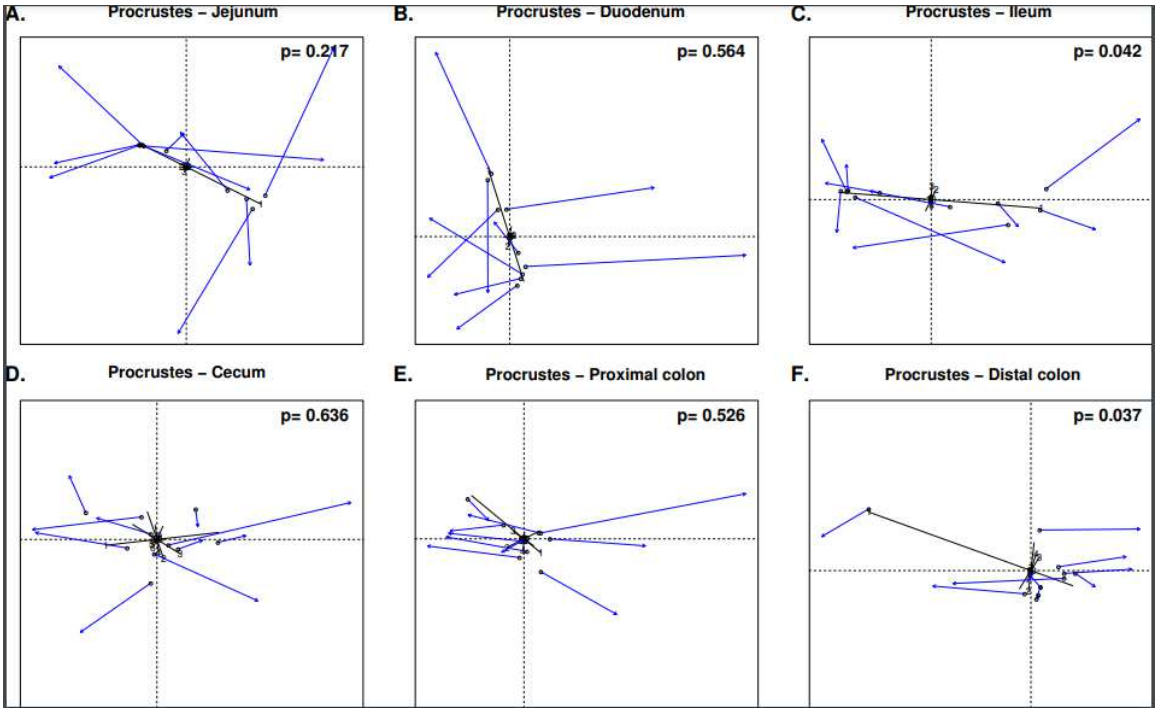

Figure 2.7: Tissue-specific Procrustes analysis.
(A) Jejunum, (B) duodenum, (C) ileum, (D) cecum, (E) proximal colon, and (F) distal colon.

2.3.3 Microbiota-metabolome interactions.

To establish a global microbiota-metabolome relationship, we performed Procrustes analysis using the *vegan* package in R (**Fig. 2.8A**). Globally, we found a significant relatedness ($P = 0.0028$) between the microbiota and the metabolome. Interestingly, this relatedness was driven by the ileum ($P = 0.042$) and distal colon ($P = 0.037$; **Fig. 2.7**). We then analyzed microbiota-metabolite relationships using Spearman's ranked correlation on the metabolites with assigned identity and abundances of bacterial genera. This includes 595 significant ($q < 0.05$, false discovery rate-adjusted *P-value*) interactions in the small intestine and 166 in the large intestine (**Table 2.1**; **Appendix Table AT4**). Additionally, we observed that the correlation network in the small intestine was more interconnected than that in the large intestine (**Fig. 2.9**). One explanation is that the large intestine harbors more bacterial species than the small intestine; thus, there could be more functional redundancies in the large intestine, with fewer correlations at greater taxonomic resolution. Interestingly, in both networks, the levels of most metabolites were correlated with only a few bacterial taxa, and such correlations tended to be in the same direction (**Table 2.1**; **Fig. 2.9**). However, a bacterial taxon tended to correlate with many metabolites in different directions. Although the current data do not demonstrate any causal relationships among the mucosal microbiota and metabolites, it nevertheless warrants further investigation.

Table 2.1: Summary of microbiota-metabolome interactions

Location	No. of metabolites with >5 bacterial interactions	No. of bacteria with >5 metabolite interactions	No. of significant interaction pairs
Small intestine	26	49	595
Large intestine	8	8	166

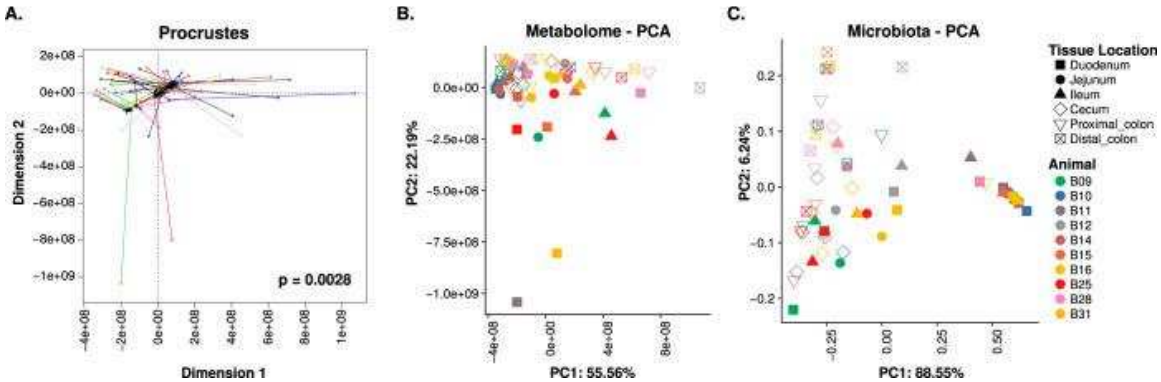

Figure 2.8: Microbiota-metabolome similarity.
(A) Procrustes analysis of the microbiota principal-component analysis (PCA) against the metabolome PCA. Longer line lengths indicate lower within-sample similarities. PCA of the tissue metabolome (B) and microbiota (C).

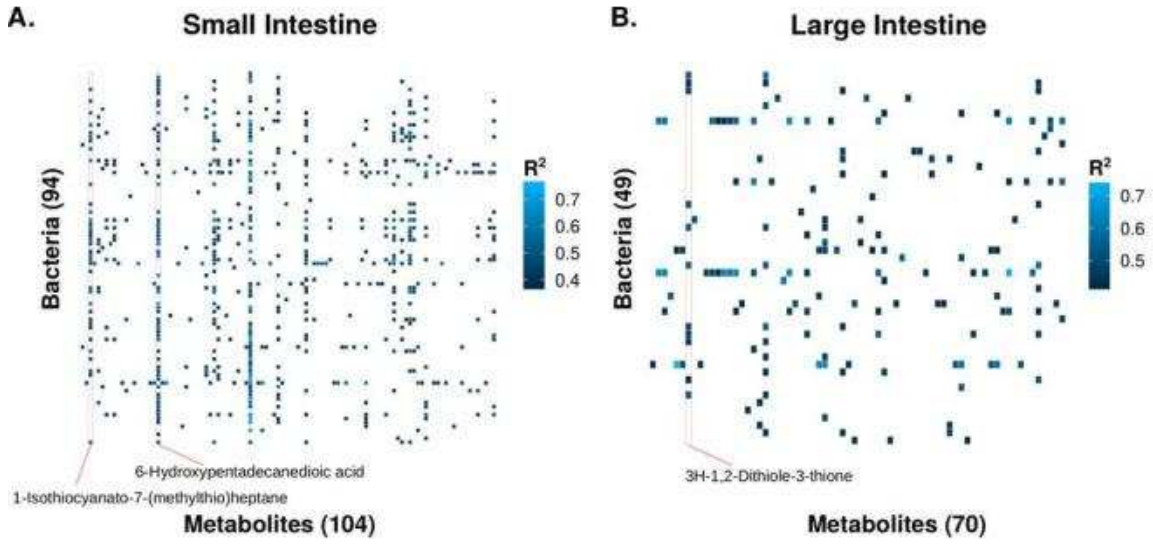

Figure 2.9: Heatmaps of significant microbiota-metabolite correlations of the small intestine (A) and large intestine (B).

2.3.4 Dietary enrichment shapes intestinal tract microbiota-metabolite interactions.

We further sought to determine the potential origin of the metabolites that were highly correlated with bacterial taxa. These metabolites corresponded to metabolites commonly found in several vegetables. 6-Hydroxypentadecanedioic acid and 1-isothiocyanate-7-(methylthio)heptane had 56 and 31 significant correlations in the small intestine, respectively (**Fig. 2.9A**). 3H-1,2-Dithiole-3-thione had 11 significant correlations in the large intestine (**Fig. 2.9B**). Surprisingly, all three compounds are commonly found in *Brassica* vegetables, which were fed to the animals as a part of the normal dietary enrichment. All bacteria significantly correlated with 6-hydroxypentadecanedioic acid in the small intestine are positively correlated with this compound (**Appendix Table AT3**). They include *Clostridium* XIVa, *Ruminococcus*, *Faecalibacterium*, and *Lactobacillus*, all of which have shown health benefits in humans [152,153]. This suggests that 6-hydroxypentadecanedioic acid may have potential prebiotic effects. Due to the potential health benefits associated with eating *Brassica* vegetables, this finding warrants additional investigation.

2.4 DISCUSSION

Currently, there is limited knowledge of the microbiota composition along different sections of the intestinal tract in either human or nonhuman primate

(NHP) samples [154,155]. Studies of human subjects usually require prior bowel preparation, which has been shown to alter the microbiota [156]. In this study, we collected tissue samples from healthy NHPs without prior bowel preparation, thus providing an unaltered view of the healthy microbiota. Previous studies have analyzed the intestinal tract microbiota compositions in mice, chickens, dogs, cows, and horses [137–141], among others. However, due to the anatomical differences, in addition to the dietary and genetic differences, these animals may have different microbiota along the intestinal tract.

Perhaps not surprisingly, the baboon microbiota composition is more similar to that observed in human intestinal tissue-associated microbiota and dissimilar to that observed in mouse fecal samples [23]. Similar to a previous human study which examined the microbiota composition using small and large intestinal biopsy samples [142], we found several microbiota differences along the intestinal tract at the phylum level. Additionally, we found lower alpha-diversity in the small intestine, while Stearns et al. did not [142]. One plausible explanation is the previous study collected biopsy samples after the patients had undergone bowel preparation, and this may affect the microbiota composition. Indeed, the fecal samples collected prior to bowel preparation had very different microbiota composition compared to the colon tissue samples.

Similar to previous reports in humans, we found variations in the microbiota composition between different NHP subjects. In addition, we found that the microbiota composition along the intestinal tract is also influenced by the host. Previous studies suggest that this variation between individuals can be

attributed to factors such as genetics, dietary preferences, and other factors [3,8,157]. In this study, differences in baboon age and weight may have further contributed to interindividual differences. The alpha-diversity differences observed along the intestinal tract sections were likely due to changes in the microbial concentration gradient, where the small intestine harbors fewer bacteria due to the high-pH environment. It is not surprising that the distal colon and cecum harbor more distinct bacterial taxa than other locations. Previous studies have shown that both the distal colon and cecum are where most bacterial fermentation takes place, although we found no discernible differences in the predicted metagenome of the microbiota [158].

In this study, we performed untargeted metabolomics on the intestinal tissues. Although we were able to identify more than 3,395 entities, we were able to assign identities to only 292 compounds. This lack of positive identification is mainly due to the lack of available databases. It is conceivable that we will be able to extract additional information from the current data in the future using improved databases, further strengthening the current research. Nevertheless, using the most current database, we found that the metabolomic profiles showed enriched cancer-related pathways in the large intestine. We think this observation suggests that the large intestinal metabolic microenvironment may better support tumor growth compared to the small intestine. This hypothesis is supported by the low incidence of small intestinal tumors in humans. However, a major caveat of the current research was our inability to distinguish between the metabolic contribution from the host and the microbiota. Thus, whether differences in

metabolite profiles between the small and large intestine were primarily driven by the microbiota is yet to be determined. A previous study comparing the tissue-level metabolome between conventional and germfree mice showed that the microbiota contribute to various metabolomic differences along the intestinal tract. However, whether this difference is due to changes in microbial metabolism or host metabolism is unclear [151]. Future studies should aim to separate metabolites originated from the host, microbiota, or food source.

Interestingly, our analysis also found that 6-hydroxypentadecanedioic acid, 1-isothiocyanato-7-(methylthio)heptane, and 3H-1,2-dithiole-3-thione, compounds commonly found in *Brassica* vegetables, were correlated with higher levels of several potentially beneficial bacteria. Notably, 3H-1,2-dithiole-3-thione has been previously shown as a potent antioxidant and potential chemopreventive agent, by targeting the transcription factor NRF2 [159]. This may suggest a potential prebiotic effect of these compounds. Moreover, these compounds show location-specific correlations with microbiota, which may suggest a potential strategy to target beneficial bacteria in different intestinal locations. One explanation for the location-specific correlations is the differences in absorption of these compounds at different locations of the intestine, which can lead to different metabolite concentrations in the intestinal lumen. However, since the current study did not include controlled feeding, we are unable to ascertain the exact role of these compounds in modulating the microbiota.

In the present study, we report the host-microbiota interactions along the healthy nonhuman primate lower gastrointestinal tract. Our study provided a

global view of the microbiota landscape of healthy NHPs. Our analysis suggests an intricate global relationship between the microbiota and metabolites along the intestinal tract. Importantly, we found that dietary components may be a means to modify microbiota composition at specific sites throughout the intestinal tract, suggesting potential targeted use as prebiotic therapeutics. Further study will be necessary to evaluate specific diet-microbiota-metabolomic interactions and the potential to use metabolites as microbiota-directed therapeutics.

2.5 PUBLICATIONS

This chapter has been modified (with permission) from the published article[160]:

160. Yuan, C., Graham, M., Staley, C. & Subramanian, S. Mucosal Microbiota and Metabolome along the Intestinal Tract Reveal a Location-Specific Relationship. *mSystems* 5, (2020). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7253361/

CHAPTER 3: INTERACTION BETWEEN HOST MICRORNAS AND THE GUT MICROBIOTA IN COLORECTAL CANCER

3.1 Introduction

The colon microenvironment hosts trillions of microbes, known as the gut microbiome. A healthy microbiome helps maintain colon microenvironment homeostasis, immune system development, gut epithelial function, and other organ functions[128–130,161,162]. Although many factors impact the composition of the gut microbiome, the overall functional profiles remain stable over time[163,164]. Nevertheless, changes in the taxonomic and functional compositions of the microbiome have been implicated in many diseases, including colorectal cancer (CRC)[23,165–167]. Although the association between microbiome alterations and disease processes has been extensively demonstrated, the directionality, as well as the mediators of the host-microbiome interaction, remains unclear.

Diet has been independently associated with both the gut microbiome and CRC. For example, the Western diet (characterized by low fiber and high protein, fat, and sugar) affects gut microbiome composition in humanized mice, whereby mice fed a Western diet have an increased abundance of *Firmicutes* and a decreased abundance of *Bacteroidetes*[168,169]. The same Western diet has also long been considered a risk factor for developing CRC[170–172]. Using an animal model of CRC, Schulz et al. demonstrated that the high-fat diet (HFD) exacerbates CRC progression; however, treating animals with antibiotics blocks HFD-induced CRC progression[173]. This suggests that diet can drive microbiome composition

change in the gut as a precursor to CRC development.

Recent studies have found that host genetic variation is correlated with microbiome composition. For example, a polymorphism near the *LCT* gene, which encodes the lactase enzyme, is associated with an abundance of *Bifidobacterium* in the gut microbiome, and microbes in the *Christensenellaceae* family were shown to be heritable, with a higher similarity between monozygotic than dizygotic twins [12,1748,9,11,175]. Another recent study investigated CRC tumors and identified a correlation between coding mutations in tumors and the composition of the microbial community in the tumor microenvironment[2]. Interestingly, in a genetic mutation model of intestinal tumors, germfree animals developed significantly fewer tumors in the small intestine [24]. Although the finding is limited to the small intestine, the trend shows that CRC development partially depends on the microbiome. In an animal model of colitis-associated CRC, Uronis et al. showed that germfree mice exhibit normal histology and do not develop tumors, compared to 62% of conventionalized mice that developed tumors (n = 13) [176]. These results support an interaction between the microbiome and host genomics that may affect tumor development.

A recent report demonstrated that fecal microRNAs (miRNAs) can shape the composition of the gut microbiome [109], indicating a mechanism by which host cells can regulate the microbial community. In CRC, several miRNAs, such as miR-182, miR-503, and mir-17~92 cluster, can regulate multiple genes and pathways and have been found to promote malignant transformation and disease

progression [41,63,70]. Interestingly, studies have also found that microbiome-derived metabolites can change host gene expression, including expression of miRNAs, in the colon [72,106]. Taken together, these results suggest a bi-directional interaction between host cells and microbes, potentially mediated through miRNA activity. However, we still know very little about the role of miRNAs in host-microbiome interactions, especially in the context of CRC. With thousands of unique miRNAs and microbial taxa present in the CRC microenvironment, it is challenging to experimentally study all possible pairwise interactions. Nevertheless, genomic characterization of both miRNA expression and microbial composition in patients with CRC can identify potential interactions between miRNAs and microbes, which can then be used as candidates for functional inspection.

Here, we establish the relationships between miRNA expression and microbiome composition in CRC patients. We sequenced small RNAs and integrated 16S rRNA gene sequencing data from both tumor and normal colon tissues from 44 patients (88 samples total). We explored the correlation between miRNAs and the microbiome through imputing the miRNA functional pathways and microbiome metabolic pathways *in silico* (**Fig. 3.1**). To our knowledge, this is the first analysis to establish a global relationship between miRNA expression and the microbiome in CRC.

3.2 Methods

3.2.1 Tissue samples.

A total of 88 matched tumor and adjacent normal tissues were collected from 44 patients by the University of Minnesota Biological Materials Procurement Network. A detailed description of sample collection was previously published [23]. Briefly, all patients provided written, informed consent. All research conformed to the Helsinki Declaration and was approved by the University of Minnesota Institutional Review Board, protocol 1310E44403. Tissue pairs were resected concurrently, rinsed with sterile water, flash-frozen in liquid nitrogen, and characterized by staff pathologists. Detailed deidentified sample metadata, including age, gender, tumor location, tumor stage, and microsatellite stability (MSS) status, are available in **Appendix Table AT5**.

3.2.2 16S rRNA sequencing and sequence analysis.

The 16S rRNA gene sequencing data were previously published[23]. Raw sequences were deposited in the NCBI Sequence Read Archive under project accession number PRJNA284355, and processed data files are available in the work of Burns et al.[23]. Briefly, total DNA was extracted from approximately 100 mg of tissue. Tissues were first physically disrupted by placing the tissue in 1 ml of QIAzol lysis solution in a 65°C ultrasonic water bath for 1 to 2 h. The efficiency of this approach was verified by observing high abundances of Gram-positive bacteria across all samples, including those from the phylum *Firmicutes*. DNA was then purified using an AllPrep nucleic acid extraction kit (Qiagen, Valencia, CA). The V5-V6 region of the 16S rRNA gene was PCR amplified with multiplexing

barcodes [177]. The bar-coded amplicons were pooled and ligated to Illumina adaptors. Sequencing was performed on a single lane on an Illumina MiSeq instrument (paired-end reads). The forward and reverse read pairs were merged using the USEARCH v7 program fastq_mergepairs, allowing stagger but no mismatches[178]. Operational taxonomic units (OTUs) were picked using the closedreference picking script in QIIME v1.7.0 and the Greengenes database (August 2013 release)[144,179,180]. The similarity threshold was set at 97%, reverse read matching was enabled, and reference-based chimera calling was disabled. The unfiltered OTU table used for the analysis is available in **Appendix Table AT6**.

3.2.3 MicroRNA sequencing.

To prepare samples for small-RNA sequencing, total RNA was extracted using an AllPrep nucleic acid extraction kit (Qiagen, Valencia, CA). RNA was quantified using the RiboGreen fluorometric assay (Thermo Fisher, Waltham, WA). RNA integrity was then measured using a model 2100 Bioanalyzer (Agilent, Santa Clara, CA). Library creation and sequencing were performed by the Mayo Clinic Genome Analysis Core. Briefly, small-RNA libraries were prepared using 1 µg of total RNA per the manufacturer's instructions for the NEBNext multiplex small-RNA kit (New England Biolabs, Ipswich, MA). After purification of the amplified cDNA constructs, the concentration and size distribution of the PCR products were determined using an Agilent (Santa Clara, CA) Bioanalyzer DNA 1000 chip and Qubit fluorometry (Invitrogen, Carlsbad, CA). Four of the cDNA constructs are pooled, and the 120-

to 160-bp miRNA library fraction is selected using Pippin Prep (Sage Science, Beverly, MA). The concentration and size distribution of the completed libraries were determined using an Agilent Bioanalyzer DNA 1000 chip and Qubit fluorometry. Sequencing was performed across 4 lanes on an Illumina HiSeq 2000 instrument (paired-end).

3.2.4 MicroRNA sequence data processing and QC.

See **Fig. 3.1** for an overview of the data analysis steps. Briefly, quality control (QC) of miRNA sequencing data was performed using FastQC before and after adaptor trimming with Trimmomatic[181]. Then, the paired-end reads were assembled using PANDAseq and aligned to the hg38 genome assembly using bowtie2[182,183]. Finally, the total mature miRNA counts were generated with HTSeq[184]. We removed 7 samples (S01, S02, S03, S36, S40, S41, and S43) due to a low number of total raw reads (fewer than 500,000 raw reads) from the analysis (**Appendix Table AT5**). A previous study showed that a number of miRNA sequencing reads as low as 500,000 provides sufficient coverage for analysis[185]. The remaining 81 samples have between 519,373 and 17,048,093 (median, 6,010,361) reads per library, with an average quality score of greater than 37 in all libraries. Between 66.79% and 96.14% (median, 83.53%) of reads passed adapter trimming (**Fig. 3.2**). Of all the reads passing adapter trimming, between 287,356 and 11,102,869 (median, 3,701,487) reads were identified as concordant pairs by PANDAseq. After being mapped to the hg38 genome, between 18,947 and

4,499,805 (median, 859,546) reads were assigned to a total of 2,588 mature miRNAs (**Fig. 3.2**). Principal-component analysis (PCA) visually shows a clear separation between tumor and normal samples (**Fig. 3.3A**), while tumor location, gender, age, total raw reads, and total mature miRNA reads do not appear to have an impact on the data (**Fig. 3.3B to F**). Similarly, PCA plots, including an additional principal component, did not detect clustering based on these factors (**Fig. 3.4**). We further performed discriminant analysis of principal components (DAPC) using the adegenet package in R, and it confirmed the existence of separate clusters for tumor and normal samples ($P < 2.2 \times 10^{-16}$) (**Fig. 3.5**) but not for gender and tumor locations ($P > 0.2$ for all comparisons) (**Fig. 3.5**)[186]. Between 283 and 1,000 (median, 670) miRNAs had coverage over 1 read, and between 134 and 599 (median, 367) miRNAs had coverage over 5 reads (**Fig. 3.3G**). Overall, the quality of our sequencing results is on par with those of previous studies and our previous observations[187].

3.2.5 MicroRNA differential expression and correlation analysis.

We identified differentially expressed (DE) miRNAs between tumor and normal samples using the DESeq2 package (1.10.1) in R (version 3.2.3)[188]. Raw miRNA counts were filtered to include miRNAs with ≥1 read in ≥80% of the samples. The remaining 392 miRNAs were then used for DESeq2 analysis. We define DE

miRNAs as showing a fold change of over 1.5, with a false-discovery rate (FDR)-adjusted *P* value (*q* value) of <0.05. We performed correlation analysis for the tumor samples using Sparse Correlations for Compositional Data (SparCC) at the genus level for bacteria and the miRNAs[189]. To increase the accuracy of estimation, we performed 20 iterations for each SparCC procedure. SparCC then performs 100 permutations to calculate the pseudo-*P* values. Significant correlations were defined as a correlation coefficient (*r*) of over 0.05 (or less than −0.05), with a pseudo-*P* value of ≤0.05. Heatmaps of the correlation were generated in R using the pheatmap package. We performed hierarchical clustering for both columns and rows with the average linkage method using Pearson's correlation. We utilized PICRUSt v1.0.0 to construct a predicted metagenome for bacteria with significant correlations with the DE miRNAs in tumor tissues[190]. Specifically, bacterial OTUs that are significantly correlated with the DE miRNAs are collapsed to the species level (L7). The predicted metagenomes are then generated by following the standard PICRUSt metagenome prediction pipeline. We included miRNAs with significant correlations with CRC-associated genera (*Fusobacterium*, *Providencia*, *Bacteroides*, *Akkermansia*, *Roseburia*, *Porphyromonas*, and *Peptostreptococcus*) to perform pathway enrichment analysis using miRPath v.3[57,191]. We generated

network visualization of miRNA-microbe using Cytoscape v3.5.1.

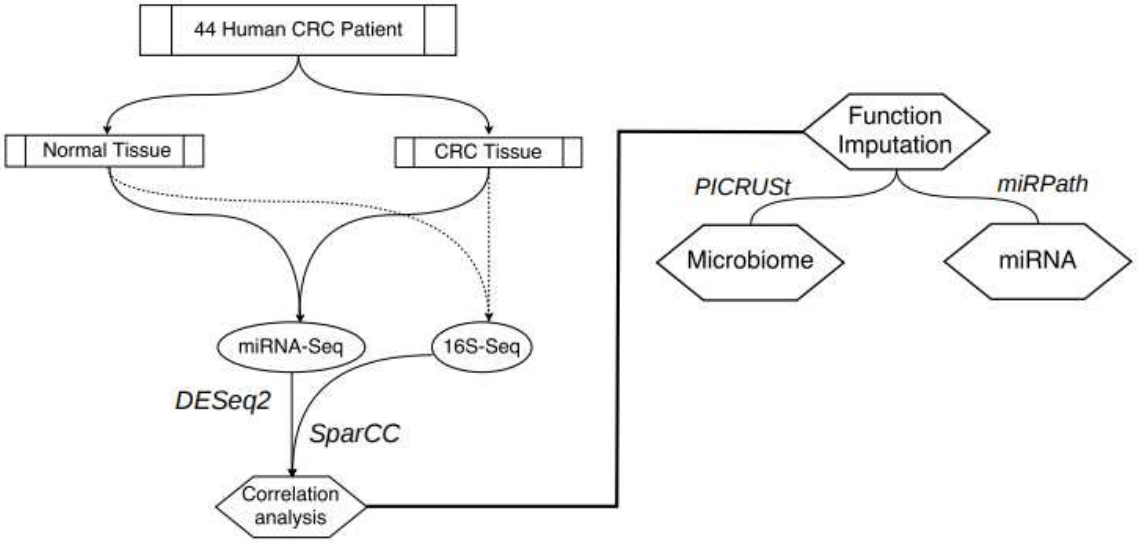

Figure 3.1: Overview of analysis.
Paired tumor and normal tissues were collected from 44 CRC patients by the University of Minnesota BioNet as a part of a previous study. The 16s rRNA gene sequencing was performed prior, as a part of the previous study. The small RNAs were sequenced on an Illumina HiSeq 2000 platform. A detailed description of the analysis was described in methods section.

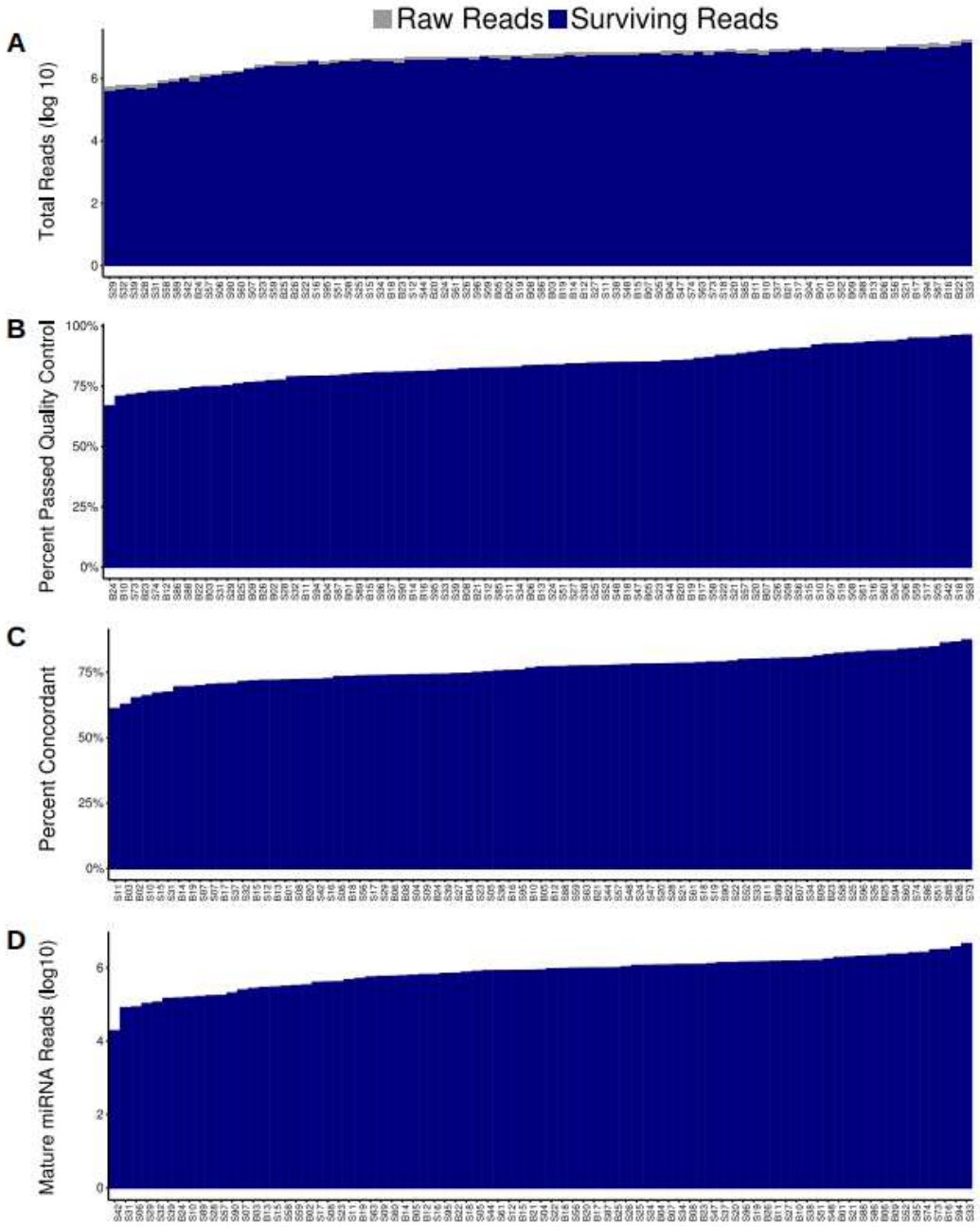

Figure 3.2: Bar plot of quality control of raw reads. a. Total number of raw reads on log10 scale per sample (grey) and percent of reads surviving quality control (blue). b. Percent of reads surviving quality control per sample. Bar plot of c. percent of concordant paired-end reads per sample and d. total number of mapped mature miRNAs per sample on log10 scale.

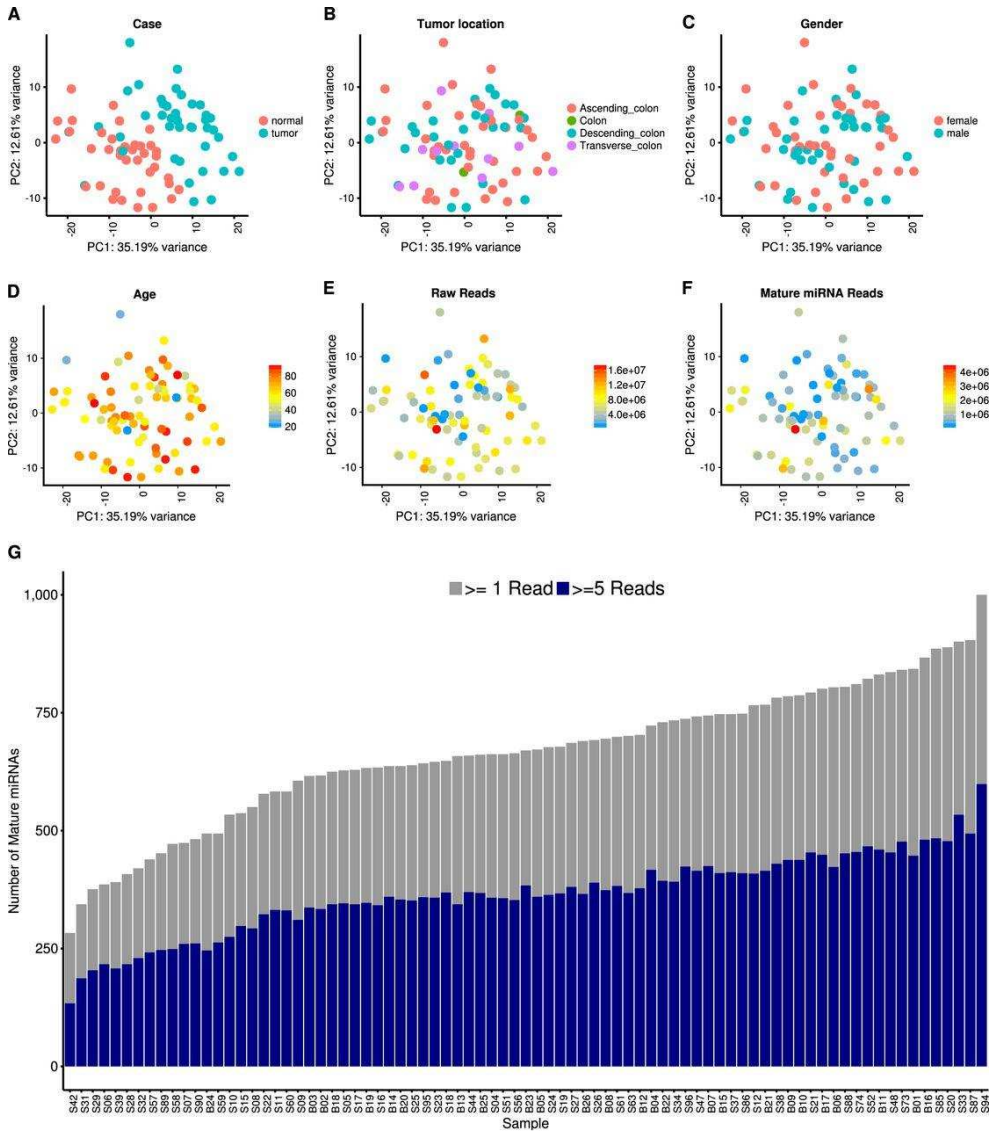

Figure 3.3: Small RNA sequencing data quality.
Principal-component analysis showing principal component 1 (PC1) on the x axis and PC2 on the y axis. Each dot is colored according to its normal/tumor status (A), tumor location (B), patient gender (C), patient age (D), raw read count (E), and mature miRNA mapped read count (F). (G) Bar plot of the numbers of mature miRNAs identified in each sample, with coverages over 1 read (gray) and over 5 reads (blue).

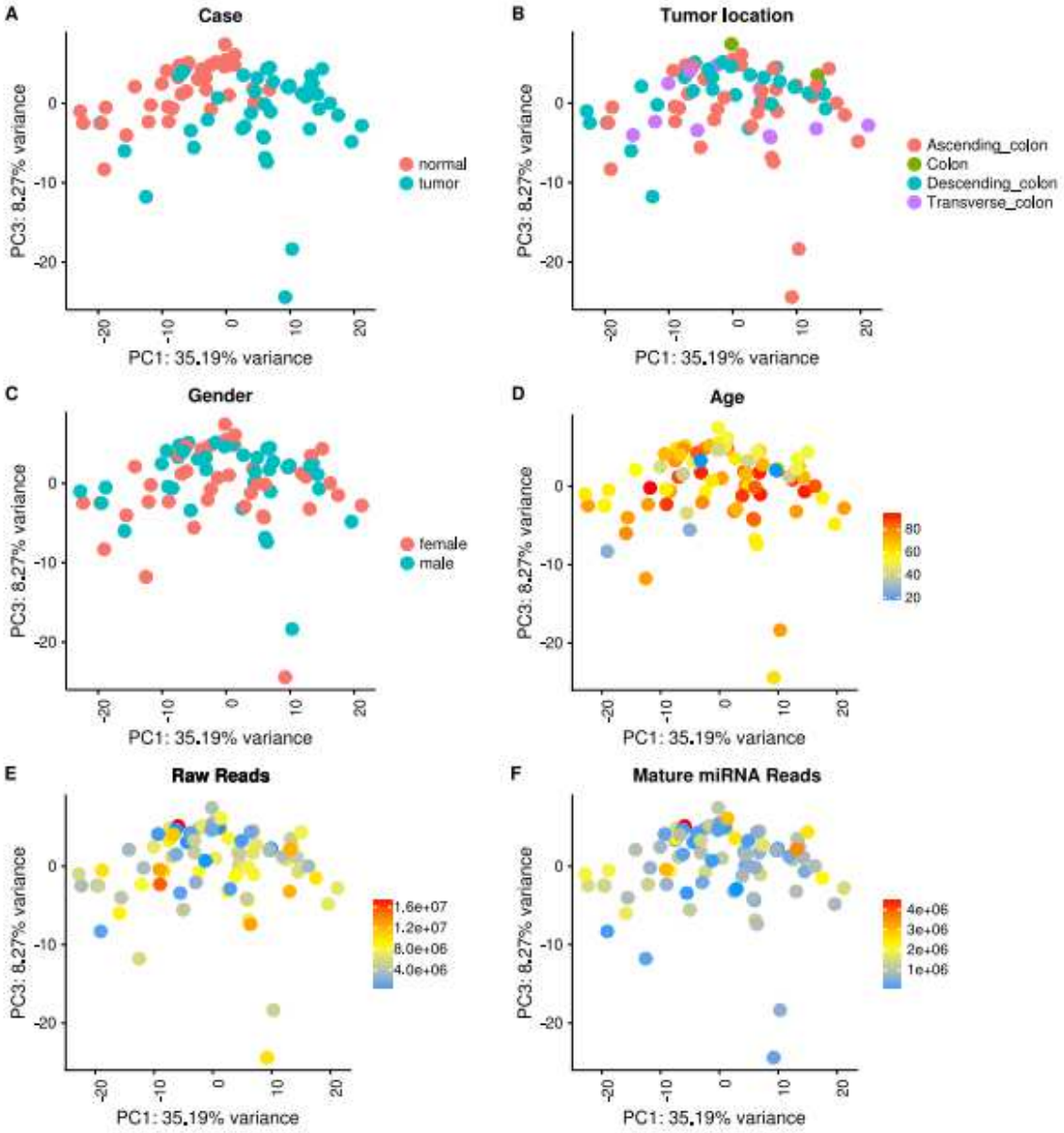

Figure 3.4: Principal Component Analysis (PCA) plot showing PC1 on x-axis and PC3 on y-axis. Each dot was colored by a. Case, b. Tumor location and c. Gender., d. Age, e. Raw reads and f. Mature miRNA Reads.

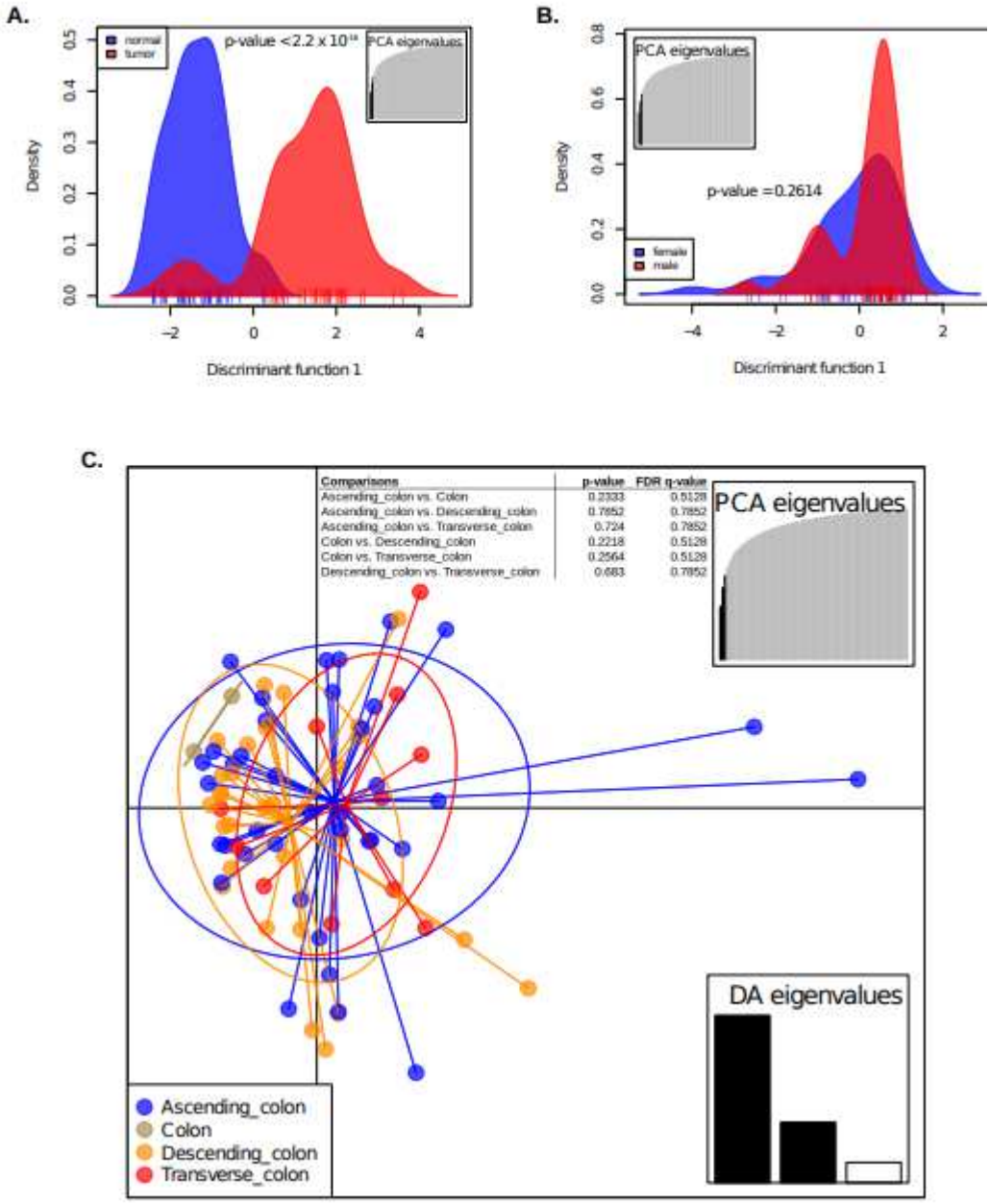

Figure 3.5: Discriminant analysis of principal components (DAPC) using the first 3 principal components. a. Case status show a clear separation between the tumor and normal groups. b. Genders and c. Tumor location do not affect the does not appear to significantly impact the miRNA expression. Statistical significances are tested using two-sample Kolmogorov-Smirnov test, multiple comparisons are adjusted using false discovery rate (FDR) method.

3.3 Results

3.3.1 MicroRNAs differentially expressed in tumor tissues.

Before performing differential expression (DE) analysis, we performed extensive quality control of the miRNA data. Our results indicate that miRNA expression is not strongly affected by tumor location, patient gender, patient age, or read coverage and show a clear clustering of miRNA data by tumor and normal samples (**Fig. 3.3**; see also Materials and Methods below). To identify small RNAs that are DE between tumor and normal samples, we performed DE analysis using DESeq2 (see Materials and Methods). A total of 76 DE miRNAs were identified, with 55 upregulated and 21 downregulated in tumor tissues compared to normal tissues (*P-value* < 0.05 after false-discovery rate [FDR] correction). A full list of DE miRNAs is available in **Table AT5** in the appendix. DE miRNAs with higher expression levels in tumor tissues include miR-182, miR-183, miR-503, and the miR-17~92 cluster miRNAs (**Fig. 3.6**; **Appendix Table AT5**), all consistent with our previous reports[59,70]. These miRNAs have all been previously shown to contribute to CRC disease progression; for example, miR-182 and miR-503 were found to cooperatively target *FBXW7* and contribute to CRC malignant transformation and progression and were also predictive of patient survival[70]. The miR-17~92 cluster regulates multiple tumor-suppressive genes in CRC and other cancers[42]. In addition, miR-1, miR-133a, and miR-448 (**Appendix Table AT5**) were observed at higher levels in normal tissues than in matched tumor tissues, also in agreement with previous reports[59,192].

3.3.2 Predicted functions of microbiome taxa correlated with DE miRNAs in tumor samples.

To investigate the relationship between individual miRNAs and the microbiome in CRC tumor samples, we performed correlation analysis using Sparse Correlations for Compositional Data (SparCC). SparCC is developed specifically to analyze compositional genomic survey data, such as 16S rRNA gene sequencing and other types of high-throughput sequencing data[189]. Hierarchical clustering revealed several clusters of significantly correlated miRNAs and bacterial taxa (**Fig. 3.7**). To further investigate the relationship between miRNAs and the microbiome in CRC, we selected bacteria significantly correlated with the DE miRNAs (**Fig. 3.8A**). The correlations clearly show a distinct pattern based on the enrichment of miRNAs, even though the correlation analysis is performed only with tumor samples. We then built a network visualizing the

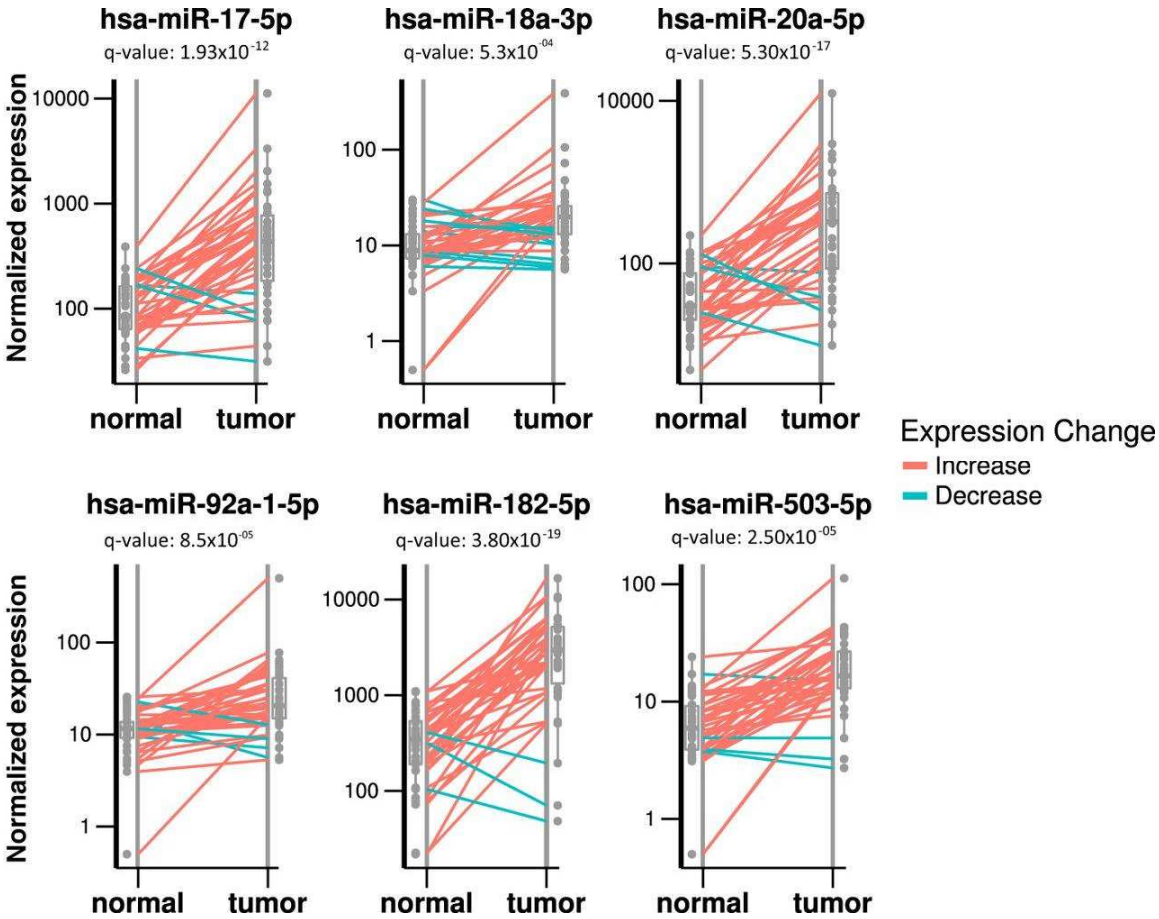

Figure 3.6: Differentially expressed miRNAs between matched normal and tumor samples. Box plot and dot plot showing differentially expressed miRNAs. Each panel represents a single miRNA with a normalized expression level on the y axis. Lines connect a normal and a tumor sample from the same individual, with red lines indicating a higher expression level in tumor tissues and green lines indicating a higher expression level in normal tissues. miR-17, -18a, -20a, -92a, -182, and -503 were found to have significantly higher expression levels in tumor tissues.

relationship between the top 9 DE miRNAs and their significantly correlated bacteria (**Fig. 3.8B** and **Fig. 3.8C**). The correlation network shows a highly interconnected relationship between these miRNAs and bacteria. Interestingly, *Blautia*, a genus previously found to have lower abundance in tumor samples, is negatively correlated with miR-20a, miR-21, miR-96, miR-182, miR-183, and miR-7974, which are all miRNAs with high expression levels in tumor tissues. *Blautia* is also positively correlated with the expression level of miR-139, which is an miRNA with high expression levels in normal tissues. Experimental validations are required to investigate the correlations.

We then analyzed the predicted functional composition of the microbiome data and investigated correlations with miRNAs (**Fig. 3.9**). We hypothesized that if miRNAs selectively affect the growth of certain bacteria, then bacteria correlated with DE miRNAs are likely to represent functional differences between tumor and normal tissues, while the uncorrelated bacteria would not. Using the PICRUSt v.1.0.0 software, we generated the predicted functional profiles of the correlated and uncorrelated bacteria by assigning pathways and enzymes using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. A total of 25 pathways have significantly altered enrichment (two-sided Wilcoxon signed-rank test with an FDR-corrected *P-value* of <0.05) (**Fig. 3.9**). Interestingly, several metabolic pathways and signaling pathways, including signal transduction, amino acid metabolism, energy metabolism, and linoleic acid metabolism, were all enriched in the uncorrelated group, suggesting increased metabolic processes in this group.

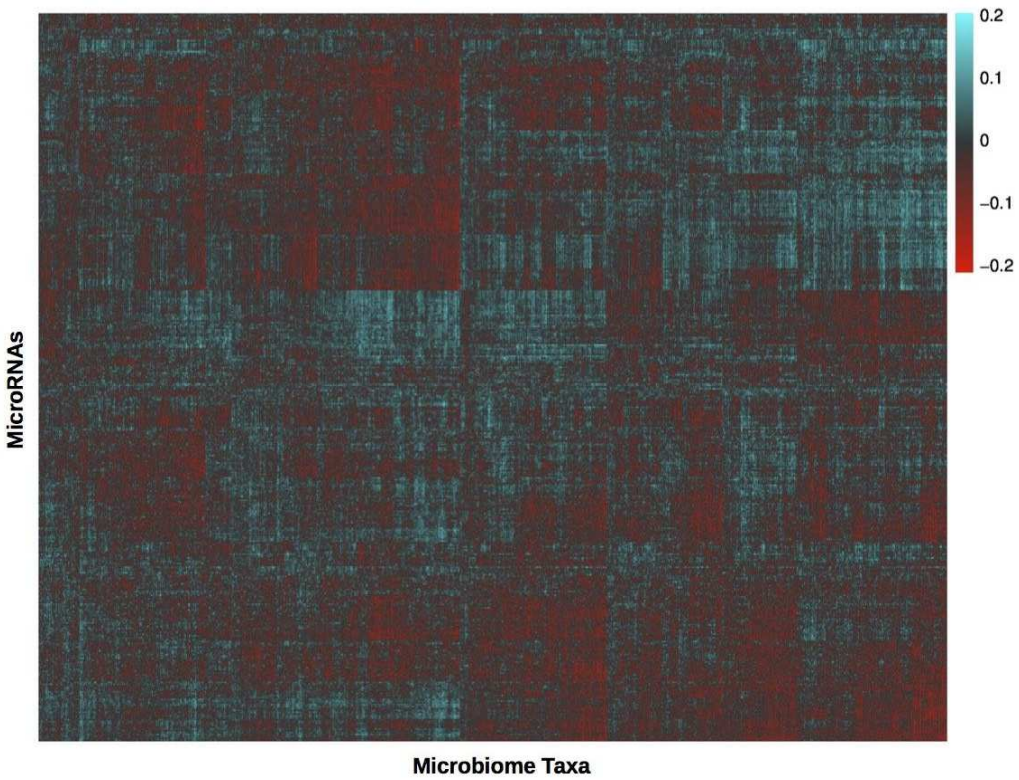

Figure 3.7: Heatmap of miRNA-microbiome correlations.
The correlations were estimated using SparCC. Bacterial taxa are shown in the column, and miRNAs are in the rows. Red indicates negative correlations, and green indicate positive correlations.

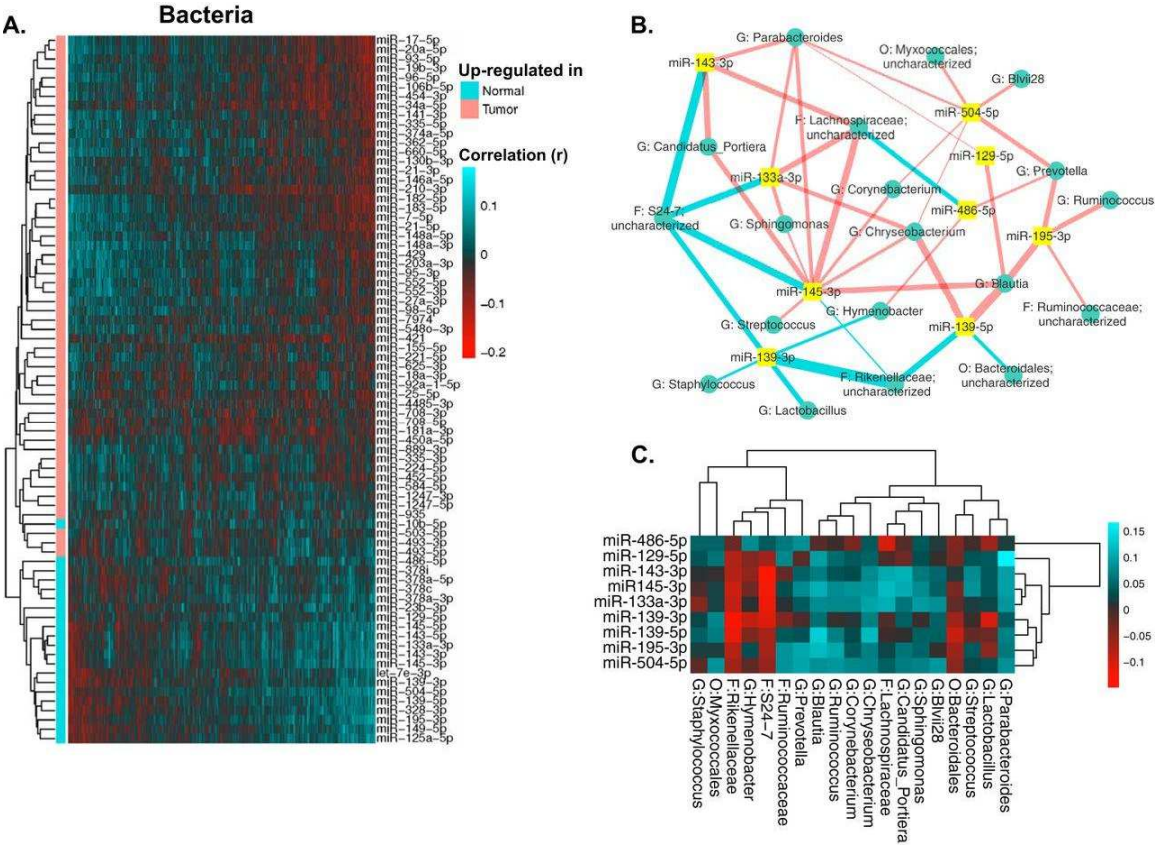

Figure 3.8: Bacteria significantly correlated with DE miRNAs.
(A) Heatmap showing bacterial genera (in columns) that were significantly correlated with the DE miRNAs (in rows). Red indicates negative correlations, and green indicates positive correlations. (B) Interaction network showing the nine most significantly DE miRNAs and their correlated bacteria (showing bacteria with a relative abundance of >0.1% and a correlation pseudo-P value of ≤0.05). Edge thickness represents the magnitude of the correlation, with blue indicating negative correlation and with red indicating positive correlation. (C) Heatmap showing the correlations displayed in panel B, with bacterial taxa in columns and miRNAs in rows. Red indicates negative correlations, and green indicates positive correlations.

For bacteria significantly correlated with DE miRNAs, however, pathways related to transporters, peptidoglycan, and terpenoid backbone biosynthesis have significant enrichment. It is worth noting that the predicted metagenome may not accurately represent the function of the microbiome; further validation using quantitative PCR or high-throughput sequencing is required.

3.3.3 Predicted functions of miRNAs correlated with CRC-associated bacteria.

To investigate the function of miRNAs correlated with CRC-associated bacteria, we focused on bacterial genera previously associated with CRC, including *Fusobacterium*, *Providencia*, *Bacteroides*, *Akkermansia*, *Roseburia*, *Porphyromonas*, and *Peptostreptococcus* [23,88,193–195]. We hypothesized that if these bacteria affect CRC through modulating miRNA expression, then miRNAs that are significantly correlated with the bacteria should show enrichment in cancer-related genes and pathways. A list of miRNAs significantly correlated with these bacteria is available in **Appendix Table AT7**. We separated these miRNAs into groups with positive correlation and negative correlation with each bacterium independently. Then, using the miRPath v.3 software, we predicted the functions of miRNAs by assigning pathways to the miRNA targets using the KEGG database (**Appendix Table AT8**). We visualized the pathways with a q value of <0.01 (modified Fisher exact test; FDR corrected) in **Fig. 3.10**.

Our results show that *Akkermansia* is the only taxon correlated with miRNAs associated with the colorectal cancer pathway. *Fusobacterium,*

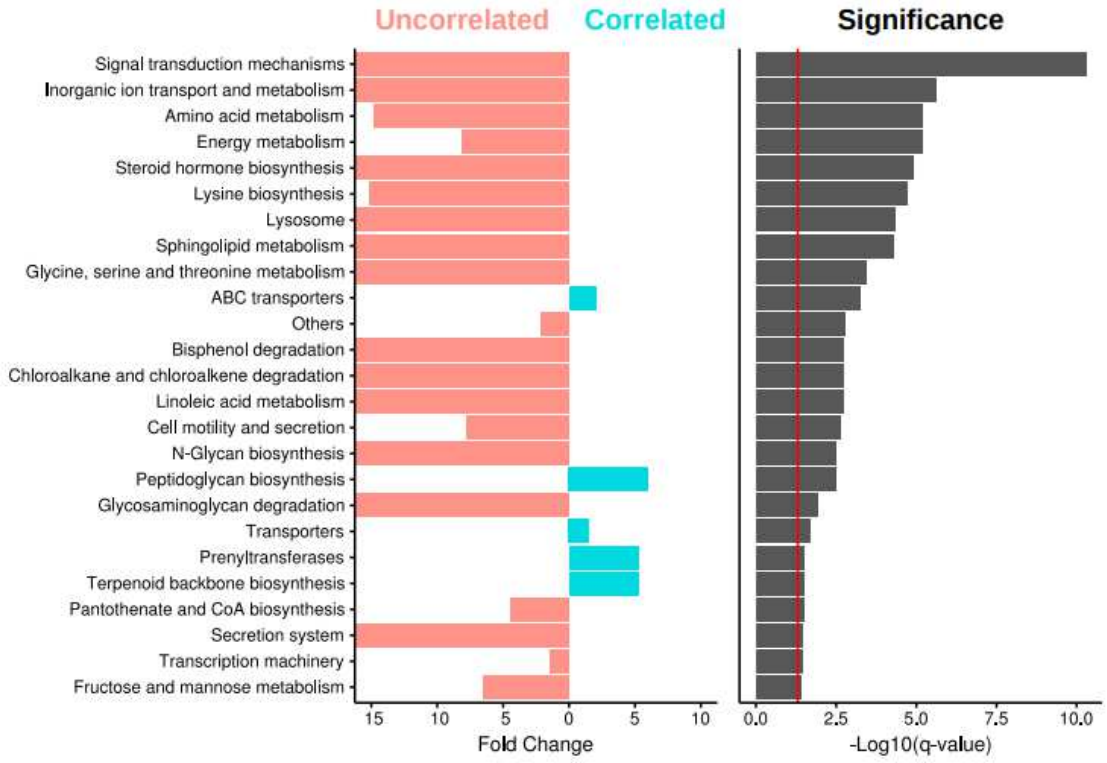

Figure 3.9: Metabolic pathway (KEGG) enrichment of microbiomes correlated and uncorrelated with DE miRNAs. The bar graph (left panel) shows the fold enrichment for each group. Red indicates correlated and blue indicates uncorrelated KEGG enrichment. FDR-corrected P values from a Wilcoxon rank-sum test (on a negative log10 scale) are shown in the right panel. The solid red line indicates a q value of 0.05

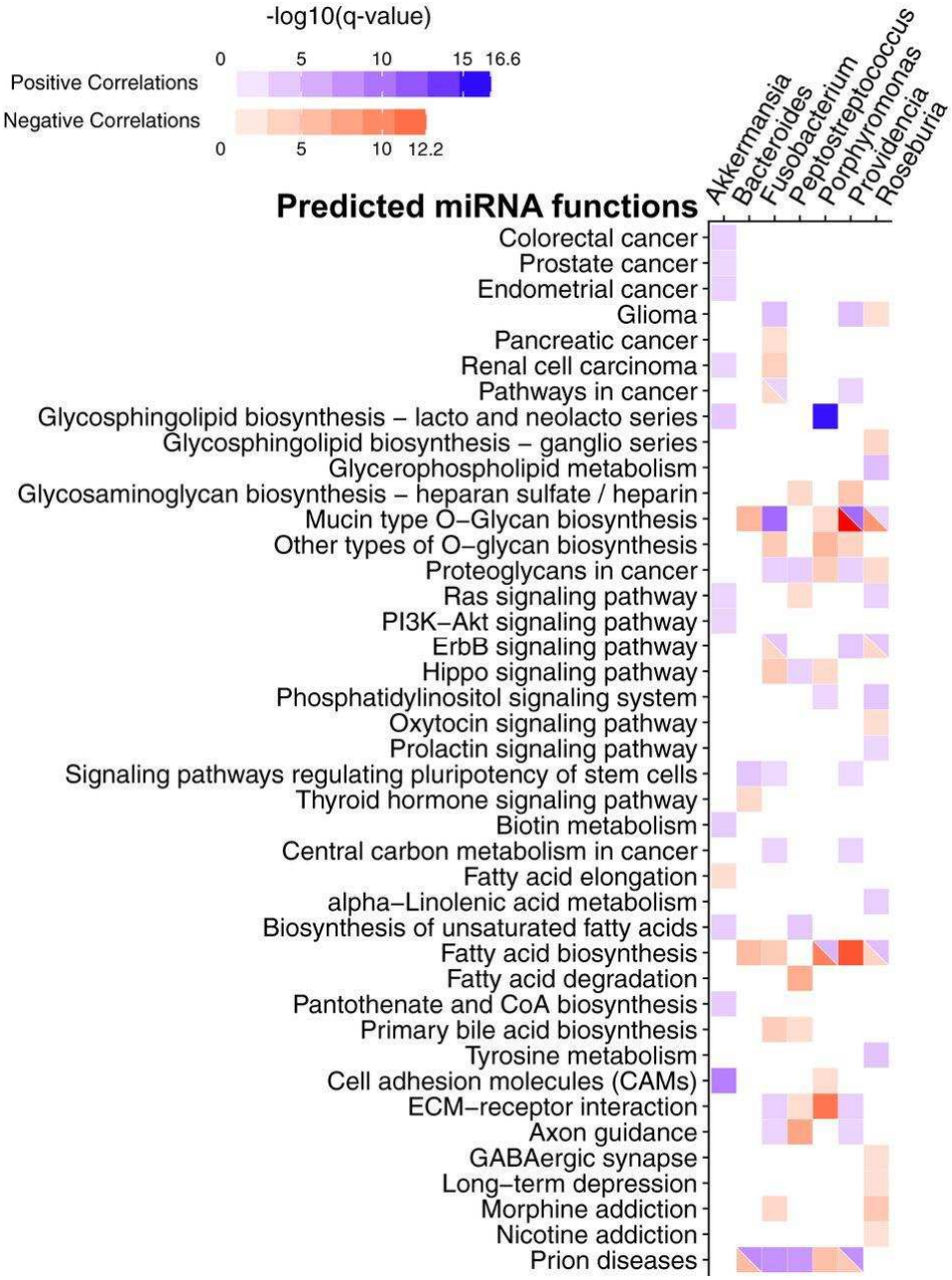

Figure 3.10: miRNA target pathways correlated with CRC-associated bacteria.

The heatmap shows the predicted pathways of miRNAs (rows) correlated with CRC-associated bacteria (columns) with a q value of <0.01 (modified Fisher exact test; FDR corrected). Positive correlations are shown in blue, and negative correlations are shown in red. The color intensity is shown in a negative log10 scale of FDR-corrected P values from a modified Fisher exact test generated by mirPath, with a darker color indicating a lower q value. CoA, coenzyme A; ECM, extracellular matrix.

Providencia and *Roseburia* correlate with miRNAs associated with cancer-related pathways, including the glioma, pancreatic cancer, and renal cell carcinoma pathways and pathways in cancer. Interestingly, glycan-related pathways, including the pathways mucin-type O-glycan biosynthesis, other O-glycan biosynthesis, glycosaminoglycan biosynthesis–heparan sulfate/heparin, and proteoglycans in cancer, have correlations with all bacterial genera analyzed, except for *Akkermansia*. This finding corresponds to a previous study showing that *Fusobacterium nucleatum* infection stimulates mucin secretion *in vitro* [196]. Additionally, *Fusobacterium nucleatum* binds to specific Gal-GalNAc, which is expressed by CRC tumors, through the Fap2 protein[19]. *Porphyromonas gingivalis* was shown to induce shedding of a proteoglycan, syndecan-1, in oral epithelial cells[197]. However, the role of the bacteria and glycan interaction is not clear in the context of CRC. Cell signaling pathways previously implicated in CRC, such as the Ras, PI3K/Akt, ErbB, and Hippo pathways, are also correlated with these bacteria[36,198–200].

3.4 Discussion

Although there is a known association between gut microbiome composition change and CRC[23,165–167], the potential mediators of this relationship remain unclear. One potential mediator is host genetics and, specifically, CRC tumor mutational profiles[24,176]. Additional evidence indicates that miRNAs can mediate host-microbiome interactions in patients with CRC[109]. Here, we presented the first

integrated analysis of miRNA expression and gut microbiome profiles in CRC patients. Our data show a highly interconnected correlation network between miRNA expression and the composition of the microbiome and support the role for miRNAs in mediating host-microbiome interactions.

Active interactions between host and the microbiome in CRC have been previously observed, leading to the proposition that pathogenic "passenger" bacteria colonizing tumor tissue might lead to exacerbated tumor progression [34]. In our analysis, we focused on potential passenger bacteria, including *Fusobacterium*, *Providencia*, *Bacteroides*, *Akkermansia*, *Roseburia*, *Porphyromonas*, and *Peptostreptococcus*. *Fusobacterium* includes several pathogenic species and is implicated in dental disease, infections, and CRC [21,201,202]. Similarly, *Providencia* has also been implicated in gastrointestinal infections [23,203–205]. The mechanism of *Fusobacterium* in promoting CRC tumorigenesis and progression has been investigated. It activates the *Wnt/β-catenin* signaling pathway through FadA protein, which binds to the E-cadherin protein on intestinal epithelial cells (IECs), thus promoting cell proliferation[21]. Several mechanisms might explain this observation. One possibility is that bacteria can infiltrate the intestinal epithelial barrier after certain pathogenic bacteria, cleaving the E-cadherin [21,206]. This might lead to an increased inflammatory response in the colon microenvironment, and the inflammation can lead to DNA damage and contribute to disease progression [21,34]. Another potential mechanism is that bacteria can directly cause mutations in IECs through virulence proteins. Several of these virulence proteins were found in *Escherichia*

coli and *Helicobacter pylori* [207,208], and results indicate that these virulence factors may be enriched in the CRC microbiome, especially in *Fusobacterium* and *Providencia*[23]. However, it is unclear whether these bacteria produce virulence proteins that can directly cause DNA damage, and further investigation is required to elucidate this mechanism.

The *Wnt/β-catenin* pathway activation by *Fusobacterium* can lead to upregulation of numerous genes related to CRC[209–211]. One such gene, *MYC*, is a transcription factor that targets multiple genes related to cell proliferation, the cell cycle, and apoptosis. The miR-17~92 cluster is a known transcriptional target of *MYC* and has oncogenic properties in several cancer types [41,42,43,44]. Interestingly, butyrate, a short-chain fatty acid produced by members of the microbiome, diminishes *MYC*-induced miR-17~92 overexpression in CRC *in vitro* through its function as a histone deacetylase inhibitor[72]. Studies of CRC have consistently found low fecal butyrate levels as well as a reduced relative abundance of butyrate-producing bacteria, such as members of the *Firmicutes* phylum[28,72,88]. One potential explanation is that, in CRC, the DE miRNAs can affect the growth of certain microbes, which eventually outcompete other species and form a biofilm on tumor tissues[109]. Indeed, our data show several enriched bacterial nutrient biosynthesis and metabolism pathways in the microbes uncorrelated with DE miRNAs, but not in the correlated group. Interestingly, pathways in bacterial cell motility and secretion are also enriched among uncorrelated bacteria, suggesting that, in addition to promoting bacterial growth, certain miRNAs may be involved in recruiting bacteria to tumor tissues. This may

also provide a possible explanation for the observed difference in alpha diversity of tumor microbiomes[23,43,44].

In our analysis of the functions of miRNAs correlated with selected bacteria known to have associations with CRC, prion disease, and morphine addiction pathways found to be enriched in our analysis do not immediately seem related to cancer (**Fig. 3.10**). Upon further investigation of miRNA target genes in these pathways, we found that several genes included in the pathways may have relevant functions in cancer. For example, mitogen-activated protein kinase (*MAPK*) is central to cell proliferation and survival, interleukin-6 (*IL6*) and interleukin-beta (*IL1β*) are cytokines involved in inflammation, protein kinase A (*PKA*) is important in regulating nutrient metabolism, Bcl-2-associated X protein (*BAX*) is a tumor suppressor gene; and prion protein (*PRNP*) are known to have a significant role in regulating immune cell function[212–214].

A recent study has suggested an additional mechanism affecting host-microbiome interactions that may promote CRC tumorigenesis and progression [19215]. Abed et al. showed that Fap2 produced by *Fusobacterium* binds to glycan produced by CRC to attach to the tumor tissue [19]. Interestingly, glycan biosynthesis pathways were enriched in targets of the miRNAs correlated with CRC-associated bacteria. The increased glycan production may increase recruitment of certain bacteria, such as *Fusobacterium*, to the tumor location. This result highlights a novel potential mechanism for miRNAs, through regulating glycan biosynthesis, to attract specific microbes to the tumor microenvironment and thus impact tumor development. Interestingly, the mucin-type O-glycan biosynthesis pathway is

enriched in miRNAs positively correlated with *Fusobacterium* but negatively correlated with *Bacteroides* and *Porphyromonas*. This suggests that these bacteria may have different mechanisms of attachment to the mucosal surface due to different abilities to bind to O-glycan[216]. Additional studies are required to test the association between *Fusobacterium*, tumorigenesis, and miRNA-driven glycan production.

It is important to note that our study uses 16S rRNA gene sequencing to characterize microbiome taxonomic composition and computationally predicted pathway composition using PICRUSt v1.0.0[190]. Although this method is widely used, metagenomic shotgun sequencing can be more accurate and informative in understanding the functional makeup of a microbial community. Similarly, to impute miRNA functional profiles, we used an *in silico* prediction method, miRPath[190,191]. While both of these methods have been rigorously tested and validated with experimental data, the results remain predictions and may not represent the real biological system[190,191]. Another limitation of our approach is that it identifies correlations and not causal relationships. Nevertheless, this approach allows us to generate a microbiome- and miRNA transcriptome-wide characterization of potential interactions, which shed light on potential new mechanisms of host-microbiome interactions.

In addition, we highlight candidates for potentially interacting host miRNAs and microbial taxa, which can be directly validated and explored in model systems [217]. For example, mouse models have been extensively used to study host-microbiome interactions in the gut[218], and studies have quantified how microbiome

colonization can modulate gene expression in the host gut[219,220]. In addition, *in vitro* approaches can be useful in dissecting the regulatory effects of the microbiome and in characterizing the effects of variation in individual taxon abundances on gene expression in host cells[221,222]. These studies can validate interactions identified in our current study and shed light on the directionality and causality.

3.5 Conclusions

Our analysis, together with evidence from previous studies, suggests that miRNAs likely mediate host-microbiome interaction in CRC. We identify potential novel mechanisms that mediate this interaction and may have a role in CRC tumorigenesis, including a possible role for miRNA-driven glycan production in the recruitment of pathogenic microbial taxa. The interactions identified here might be a direct target for developing therapeutic strategies that can benefit CRC patients. Follow-up studies using model systems are warranted to assess the causal role of individual microbes and miRNAs in CRC.

3.6 PUBLICATIONS

This chapter has been modified (with permission) from the published article[3]:

3. Yuan, C., Burns, M. B., Subramanian, S. & Blekhman, R. Interaction between Host MicroRNAs and the Gut Microbiota in Colorectal Cancer. *mSystems* 3, e00205-17 (2018).
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5954203/

CHAPTER 4: MICROBIOTA – METABOLITE – IMMUNE CELL INTERACTIONS IN COLORECTAL CANCER

4.1 Introduction

The gut microbiota actively metabolizes undigested food and substances shed from the intestinal cells, thereby generating energy and sending vital nutrients back to the host[28]. Without the microbiota, the colon's epithelial cells will undergo autophagy and will fail to maintain their structure[29]. In the normal colon, epithelial cells primarily use butyrate as energy. Tumor cells, however, require a large amount of glucose as their energy source to sustain growth, creating a large amount of lactate as the product in the tumor microenvironment. In addition, to support the formation of new cell membranes, the tumor has an increased need for lipid and amino acids biogenesis. The change in the energy source preferred by proliferating tumor cells profoundly alters the nutrient composition of the tumor microenvironment (TME). The TME is immensely complex that includes the microbiota, host epithelium, and infiltrating immune cells in the colonic mucosa. The microbiota helps ferment host dietary products and supply essential metabolites to the host. Alterations in the microbial profiles in the TME can negatively affect the activity and function of tumor-infiltrating immune cells, including macrophages, and T cells. Signaling interactions between these species at the GI tract maintain host-microbiota physiology and function[16-18]. Recent analysis in metagenomics and metabolomics have yielded associations of the microbiota with numerous disease pathologies including CRC[19-21,23,165,166,223]. Yet the directionality and the mediators between CRC and

dysbiosis remain unclear[3]. Given the drastic changes in the nutrient composition of the tumor microenvironment and the role of the microbiota in metabolism, there is undoubtedly a metabolic interaction between the tumor and its microbiota. Investigating the metabolic secretions from the bacteria may produce a more standardized and reproducible approach to identify candidates that can mechanistically alter immune cell phenotype and function within the CRC TME. Modulatory factors secreted by the microbiota include indole metabolites, Short Chain Fatty Acids (SCFAs), bile acids, and amino acids that have been implicated in shaping the presence and function of infiltrating immune cells.

To investigate microbial metabolite contribution to altered immune cell infiltrate we employed an integrated multi-omics approach of untreated primary tumor and tumor matched adjacent normal tissues. We performed 16S rRNA sequencing, RNA sequencing, and untargeted metabolomics to address this looming question. This comprehensive data set and data integration yielded a robust characterization of the TME within our cohort.

4.2 Materials and Methods

4.2.1 Sample collection and tissue preparation

A detailed sample collection method was previously described[224]. A total of 24 treatment naïve human CRC biospecimens along with matched adjacent normal tissue were collected through the Biological Materials Procurement

Network (BioNet) of the University of Minnesota and the Cooperative Human Tissue Network under an institutional review board (IRB)-approved protocol. All freshly resected specimens were transported in ice-cold RMPI-1640 medium and immediately flash-frozen in liquid nitrogen in the laboratory. The frozen tissues were stored in -80C freezer until further processing for RNA-seq, 16S-seq, and metabolomics analysis.

Samples were prepared using the same method as described in the materials and methods section of **Chapter 2**. Briefly, the tissues were ground using a mortar and pestle in liquid nitrogen. The ground tissues were then used for DNA, RNA, and metabolites preparation.

4.2.2 RNA and 16S gene Sequencing and data analysis

The DNA and RNA were isolated using the Qiagen Allprep kit (Qiagen) with QiaCube machine following the manufacture's protocol. The University of Minnesota Genomics Core performed library creation and sequencing. The RNA sequencing method was described in detail in a previous study [225]. Briefly, 1 ug of total RNA was used to create each sequencing library using the StrandedRNA Pico Mammlian LP (TakaraBio) and was sequenced on a NovaSeq (Illumina) instrument with the 150-base pair paired-end mode. The FASTQ files were analyzed using a customized pipeline (gopher-pipelines, https://bitbucker.org/jgarbe/gopher-pipelines/overview) developed and maintained by the Minnesota Supercomputing Institute. Differential gene

expression analysis was then performed in R, version 3.4.3, using the edgeR package (https://biocoductor.org/packages/release/bioc/html/edgeR.html).[25] We performed an in-silico immune cell infiltration analysis using the CIBERSORTx algorithm (https://cibersortx.stanford.edu/) following the recommended settings.[26]

For 16S gene sequencing, the V3-V5 region of the 16S rRNA gene was amplified as described in **Chapter 2**. Sequencing was performed on a single lane of a MiSeq sequencer (Illumina) using paired-end mode. The sequencing results were analyzed using the IM-TORNADO2 pipeline[143]. Alpha- and beta-diversity metrics were analyzed using QIIME v1.9.1[144].

4.2.3 Untargeted Metabolomics and Data Analysis

The metabolite extraction and untargeted metabolomics were performed using the same protocol described in detail in **Chapter 2**. The data were processed using Progenesis QI software (Thermo). The software first aligns all the features obtained in all the runs and then assigns intensity measure for features found in all the runs. The raw data were further processed by filtering for the fidelity of individual feature detection using the quality control samples. Only features with a coefficient of variation (CV) of less that 10% overall quality control samples were accepted. Features showing high intensity in background samples relative to quality control samples and features not present in at least 67% of all samples were removed from analysis per the U.S. Food and Drug Administration

recommendation. Each feature is uniquely identified with the mass-to-charge ratio (*m/z)* and the elution time form the column. Features were then assigned to metabolites identified by searching the Human Metabolome DataBase (HMDB) and using databases developed by the University of Minnesota. Co-pathway analysis was performed using MetaboAnalyst 5.0 (https://www.metaboanalyst.ca/).

4.2.4 Correlation analysis.

All correlation analyses were performed in R v3.4.4 using the Spearman's ranked correlation test with false-discovery rate (FDR) adjustment[149]. The protocol was described in detail in **Chapter 2**. Briefly pair-wise Spearman's ranked correlation test was calculated using the *cor. test* function. The *P* values were then adjusted using the *p.adjust* function before filtering for significant correlations.

4.3 Results

4.3.1 Differential analysis between tumor and normal tissues

We observed significant differences between the tumor and normal tissues between their gene expression profiles ($p < 0.03$; **Fig. 4.1A, B; Fig. 4.2A, B**). There were a total of 2,969 differentially expressed genes (fold change > 2

and FDR-adjusted $p < 0.05$). Of those genes, 1,283 were upregulated in the tumor tissues and 1,686 upregulated in the normal tissues.

The overall metabolome differences were not statistically significant between tumor and normal (**Fig. 4.1C, D; Fig. 4.2C, D**). In total, we have identified 38 statistically significant differentially abundant metabolites, of which 3 have lower levels and 35 have higher levels in tumor tissues (FDR-adjusted $p << 0.1$; **Table 4.1**).

Table 4.1: Differentially abundant metabolites between tumor and normal tissues.

HMDB ID	Normal	Tumor	FDR-adjusted p-value
HMDB0000162	19.83	20.99	5.45E-05
HMDB0000721	14.26	15.74	0.0001244
HMDB0000182	17.82	18.44	0.0001498
HMDB0000070	17.28	17.94	0.0001797
HMDB0000300	16.41	18.71	0.0003662
HMDB0000684	12.81	14.69	0.0006013
HMDB0000157	22.31	23.57	0.0007016
HMDB0010383	17.55	19.43	0.0009547
HMDB0008002	11.82	17.41	0.001994
HMDB0011737	17.19	17.95	0.001994
HMDB0000159	18.61	19.45	0.002986
HMDB0010379	16.76	18.53	0.002986
HMDB0034365	12.54	15.01	0.003365
HMDB0000687	21.89	22.61	0.004354
HMDB0000063	14.83	13.60	0.004919
HMDB0000167	17.22	17.85	0.005547
HMDB0000158	17.59	18.19	0.008007
HMDB0000251	20.71	21.74	0.008968
HMDB0000123	14.18	14.94	0.01003
HMDB0001397	20.25	21.14	0.01003
HMDB0000883	19.60	20.27	0.01003
HMDB0000875	15.81	16.60	0.01003
HMDB0000292	18.42	19.91	0.01090
HMDB0000156	18.99	19.64	0.01215
HMDB0000267	18.33	17.56	0.02122
HMDB0005060	16.56	18.51	0.02346
HMDB0000283	18.11	18.71	0.02589
HMDB0031240	11.45	9.67	0.02634

HMDB0010391	18.13	19.63	0.02827
HMDB0002231	17.31	18.84	0.04243
HMDB0000939	17.80	19.14	0.04243
HMDB0000092	10.32	11.71	0.05087
HMDB0002815	22.84	24.00	0.06755
HMDB0009783	23.74	24.97	0.06755
HMDB0010392	17.28	18.56	0.07286
HMDB0000172	21.94	22.57	0.07928
HMDB0000273	8.82	10.26	0.08618
HMDB0000630	12.27	13.14	0.09357

The overall microbiota profile is different between tumor and normally based on unweighted Unifrac distance ($p<0.05$; **Fig. 4.1E, F; Fig. 4.2E, F**), however, no significant differences are observed based on weighted Unifrac distance ($p<0.8$). Overall, there was 11 genre significantly different between the tumors (**Table 4.2**; $p<0.05$). While some groups have reported larger microbiota differences between the tumor and adjacent normal tissue, others have shown limited differences[226–229s]. This could potentially be due to differences in surgical sample collection, the relative distance between the normal and tumor tissues as well as the location of the tumor.

Table 4.2: Differentially abundant microbes between tumor and normal tissues.

ID	Normal	Tumor	p-value
Propionibacterium	0.03239845	0.00634185	0.005
Roseburia	0.00099343	0.00266521	0.013
Corynebacterium	0.00146358	0.00014692	0.026
Pseudomonas	0.01087837	0.00089081	0.029
Aquabacterium	0	0.00057164	0.036
Bibersteinia	0.00397880	0.00144829	0.044
Lactonifactor	0.00013136	0	0.045
Oribacterium	0.00032415	0	0.045
Acidocella	0.00235101	0	0.045
Clostridium_XI	0.00421386	0.020029928	0.046
Streptophyta	0.07586353	0.067695203	0.049

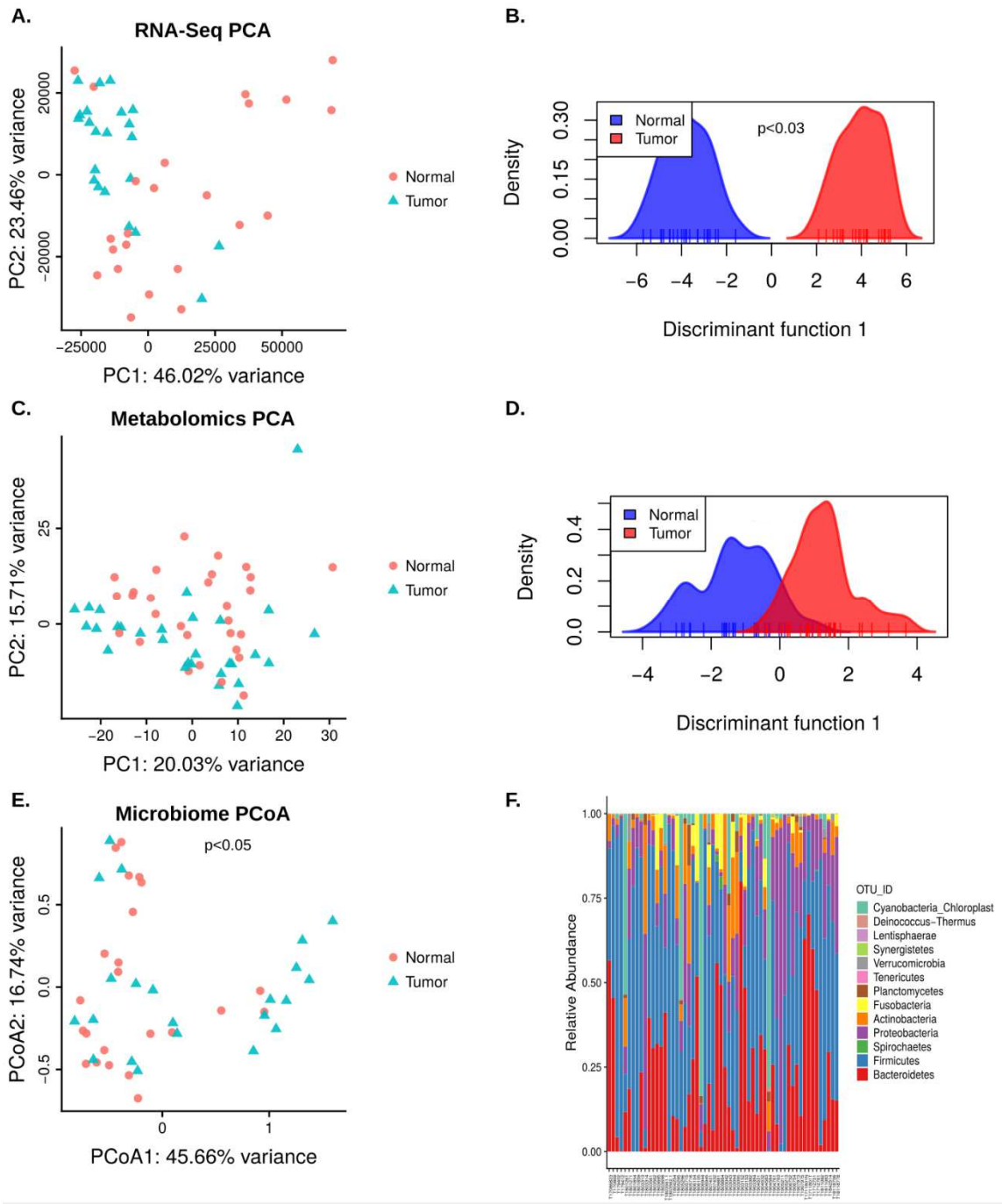

Figure 4.1: Differential analysis between tumor and normal tissues. a), Principal component analysis of RNA-Seq. b) Discriminant analysis of principal components of RNA-Seq data. c) Principal component analysis of untargeted metabolome data. d) Discriminant analysis of principal components of untargeted metabolome data. e) Principal coordinate analysis of unweighted unifrac distance of the microbiota data. f) Phyla level relative abundance.

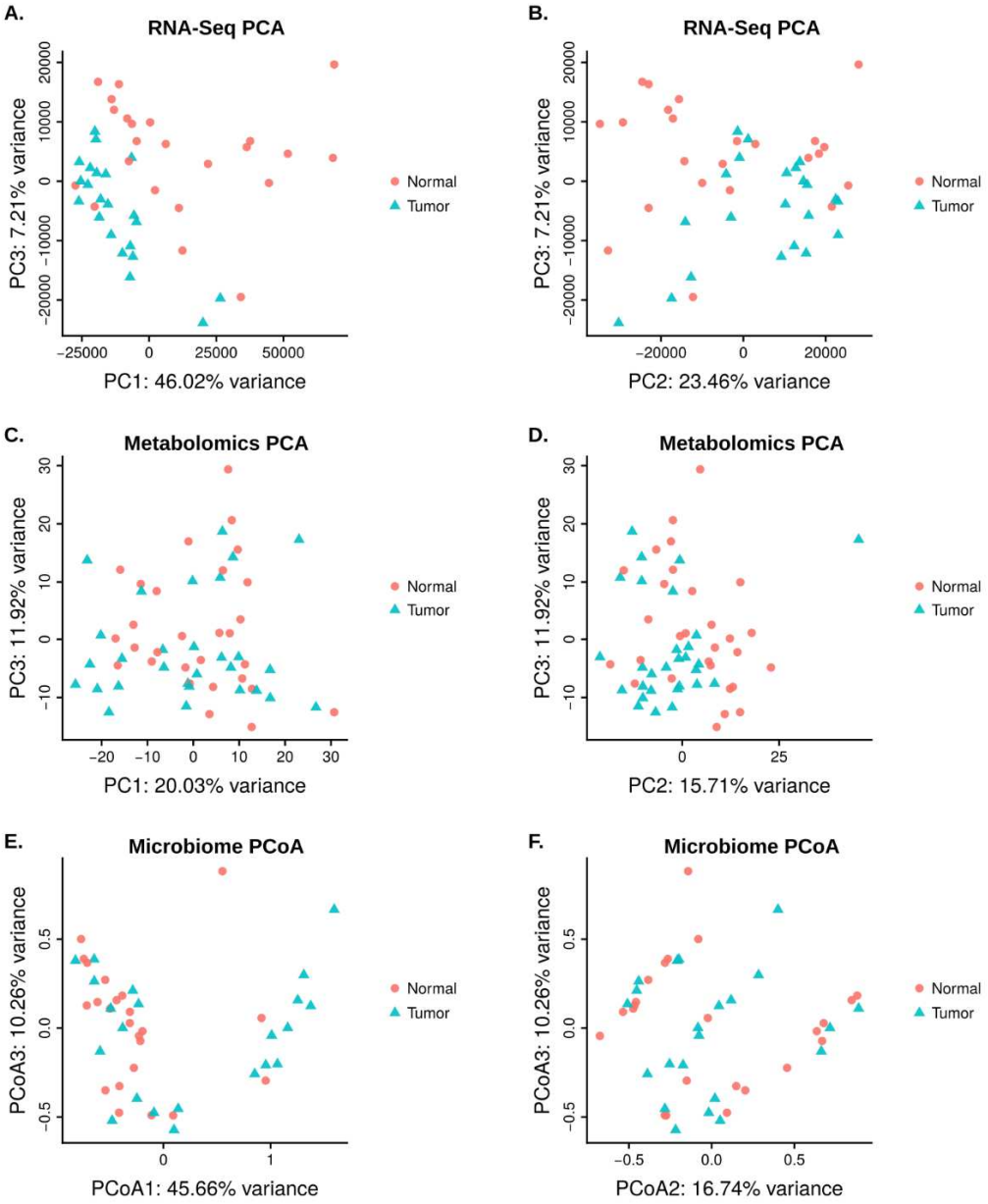

Figure 4.2: Principal component/coordinate analysis between tumor and normal tissues.
Principal component analysis of RNA-Seq data showing a) PC1 and PC3. b) PC2 and PC3. Principal component analysis of untargeted metabolome data showing c) PC1 and PC3. d) PC2 and PC3. d) Principal coordinate analysis of unweighted unifrac distance of the microbiota data showing e) PC1 and PC3. f) PC2 and PC3.

4.3.2 Pathway analysis

We then sought to investigate the pathway enrichment with changes in the metabolome. We found that 3 pathways were significantly enriched, including Aminoacyl-tRNA biosynthesis ($p < 1e-8$), Valine, Leucine, and Isoleucine Biosynthesis ($p < 0.0005$), phenylalanine, tyrosine, and tryptophan biosynthesis ($p < 0.1$). Curiously, all three pathways are related to amino acid biosynthesis, and the amino acids involved, except for tyrosine, are all essential amino acids, meaning humans cannot synthesize those amino acids. While tyrosine is a conditionally essential amino acid, which can be synthesized from phenylalanine in humans. We found that all of these amino acids are present at a higher concentration in the tumor tissues (**Table 4.3**). In **Chapter 2**, we found a higher concentration of amino acids in the Baboon small intestines (**Appendix Table AT3**), this is consistent with our knowledge that amino acid absorption occurs in the small intestines. Considering humans cannot synthesize these amino acids, we postulate that tumor microbiota could be involved in synthesizing these amino acids that are required for cellular functions.

Table 4.3: Select amino acid levels in tumor and normal tissues.

HMDB ID	Normal	Tumor	FDR-adjusted p-value
Valine (HMDB0000883)	19.60	20.27	0.01003
Leucine (HMDB0000687)	21.89	22.61	0.004354
Isoleucine (HMDB0000172)	21.94	22.57	0.07928
Phenylalanine (HMDB0000159)	18.61	19.45	0.002986
Tyrosine (HMDB0000158)	19.83	20.99	5.45E-05
Tryptophan (HMDB0000929)	14.26	15.74	0.0001244

Next, we performed an integrative Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis by combining DEGs and DE metabolites in the same analysis (**Figure 4.3-8**). We identified 5 significantly enriched pathways, including Nitrogen metabolism ($p<0.0001$; **Figure 4.3**), purine metabolism ($p<0.005$; **Figure 4.4**), phenylalanine metabolism pathway ($p<0.06$; **Figure 4.5, Figure 4.8**), Mucin type O-glycan biosynthesis ($p<0.1$; **Figure 4.6**), glutathione metabolism ($p<0.1$; **Figure 4.7**). Interestingly, of the 5 pathways, 4 are related to amino acid metabolism. Although the Mucin type O-glycan biosynthesis pathway was not related to amino acid metabolism, it was identified in **chapter 3** to be associated with *Fusobacterium,* a known carcinogenic bacterium[196].

4.3.3 CRC Interaction networks surrounding essential amino acids

Based on the differential expression and pathway analysis, we then focused our analysis on the essential amino acids. We mapped the interaction network around these differentially abundant essential amino acids, including microbiota as well as immune cell fractions. We decided to include imputed immune cell fractions due to previous studies indicating the potential role of the essential amino acids in immune cell regulations (**Figure 4.9**). The network analysis shows that CD4 T resting memory T cells had the most correlations with the microbes and metabolite isoleucine (HMDB0000172). Leucine

(HMDB0000687) is positively correlated with both M2 Macrophages and resting Mast cells. Tryptophan (HMDB0000929) is negatively correlated with *Aquabacterium*, *Coprococcus*, and activated memory CD4 T Cells. Valine (HMDB0000883) is positively correlated with *Serpens* and *Oribacterium*. Tyrosine (HMDB0000158) is positively correlated with *Micrococcus* and *Acidocella*. Phenylalanine (HMDB0000159) is negatively correlated with *Corynebacterium* and *Bibersteinia*. Notably, we also found both *Corynebacterium* and *Bibersteinia* to have higher relative abundance in the tumor tissues.

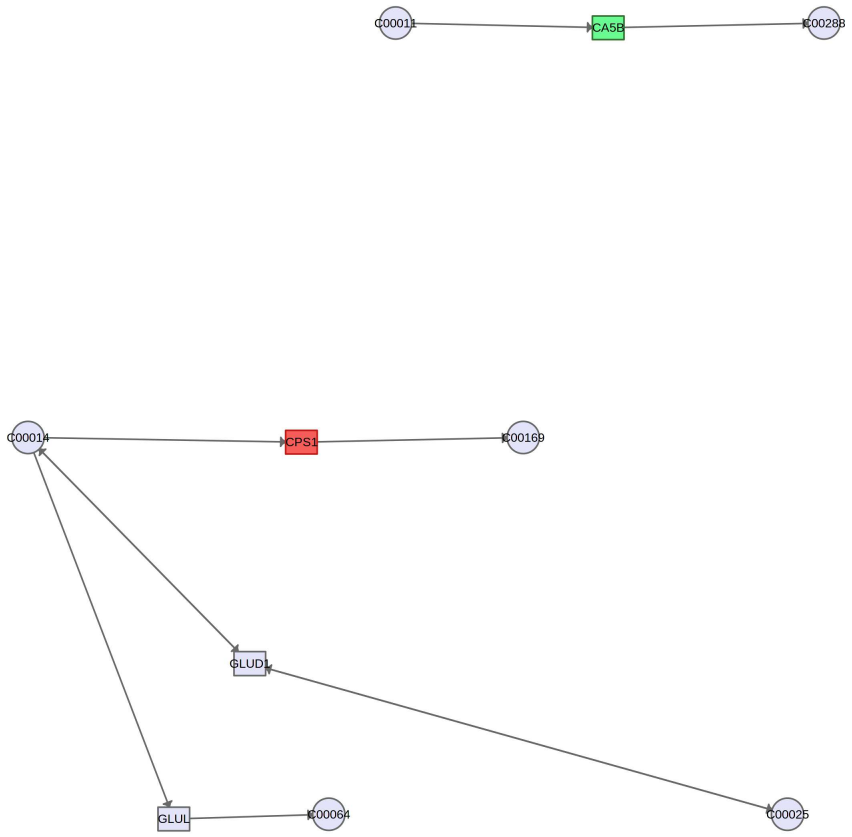

Figure 4.3: KEGG metabolic pathway co-enrichment analysis. Nitrogen metabolism.
Red indicates higher expression level (genes) or concentration (metabolites) in the tumor tissue and green indicates lower levels.

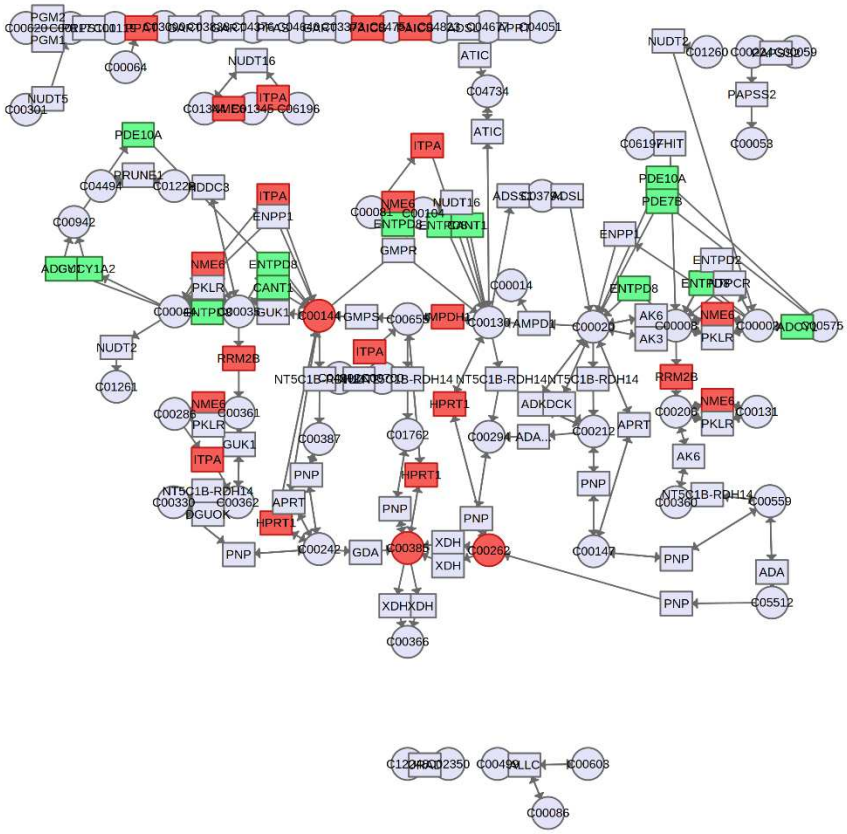

Figure 4.4: KEGG metabolic pathway co-enrichment analysis. Purine metabolism.
Red indicates higher expression level (genes) or concentration (metabolites) in the tumor tissue and green indicates lower levels.

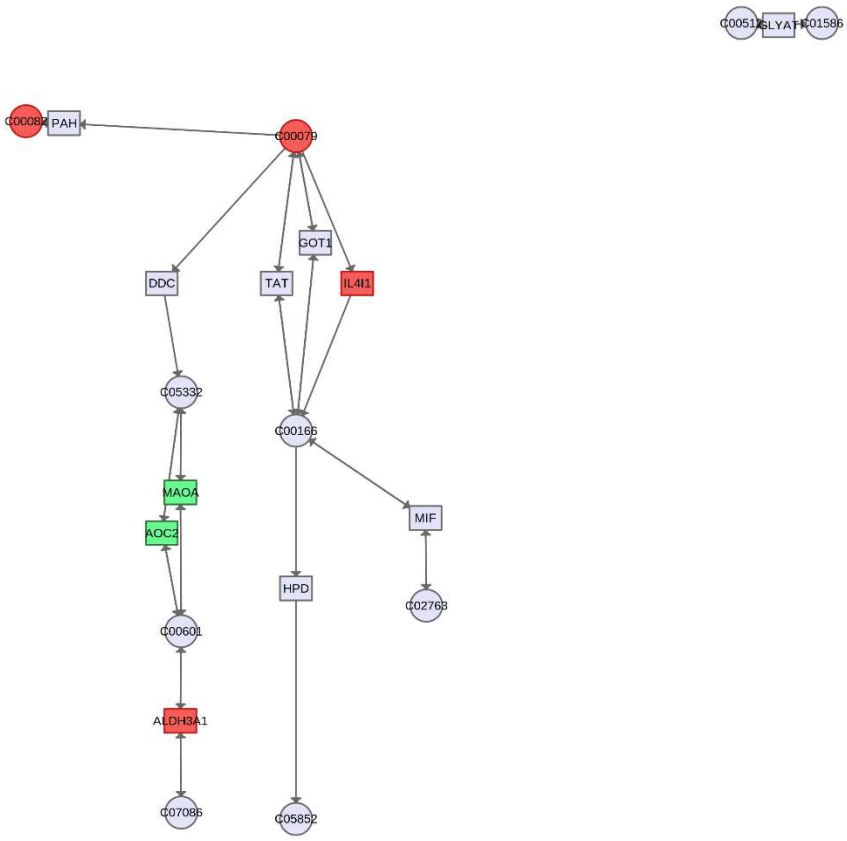

Figure 4.5: KEGG metabolic pathway co-enrichment analysis. Phenylalanine metabolism.
Red indicates higher expression level (genes) or concentration (metabolites) in the tumor tissue and green indicates lower levels.

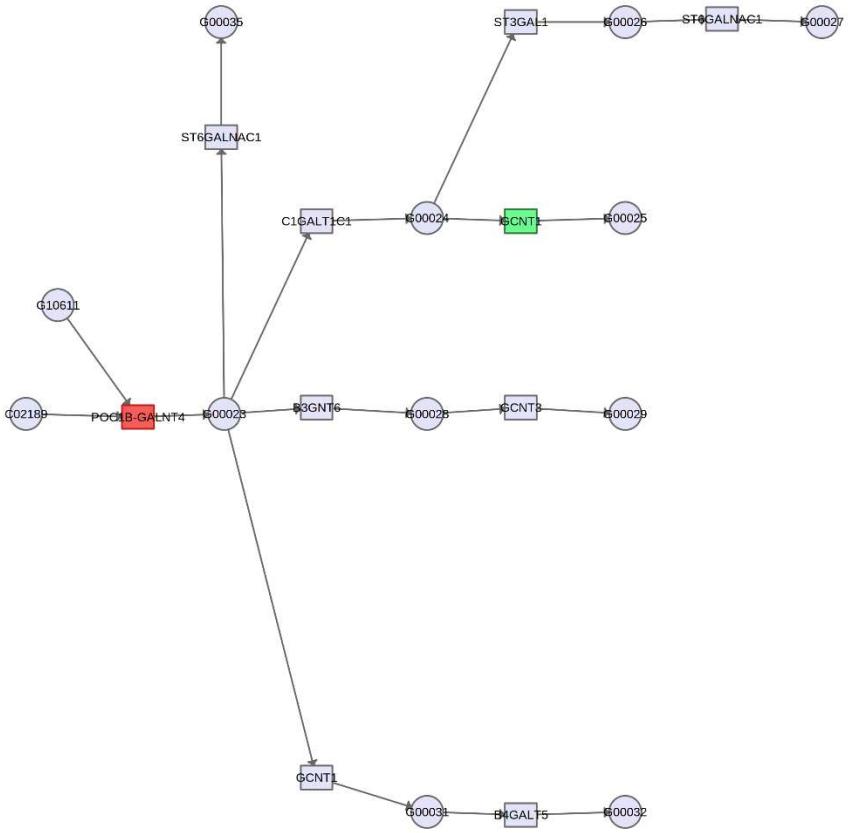

Figure 4.6: KEGG metabolic pathway co-enrichment analysis. Mucin type O-glycan biosynthesis.
Red indicates higher expression level (genes) or concentration (metabolites) in the tumor tissue and green indicates lower levels.

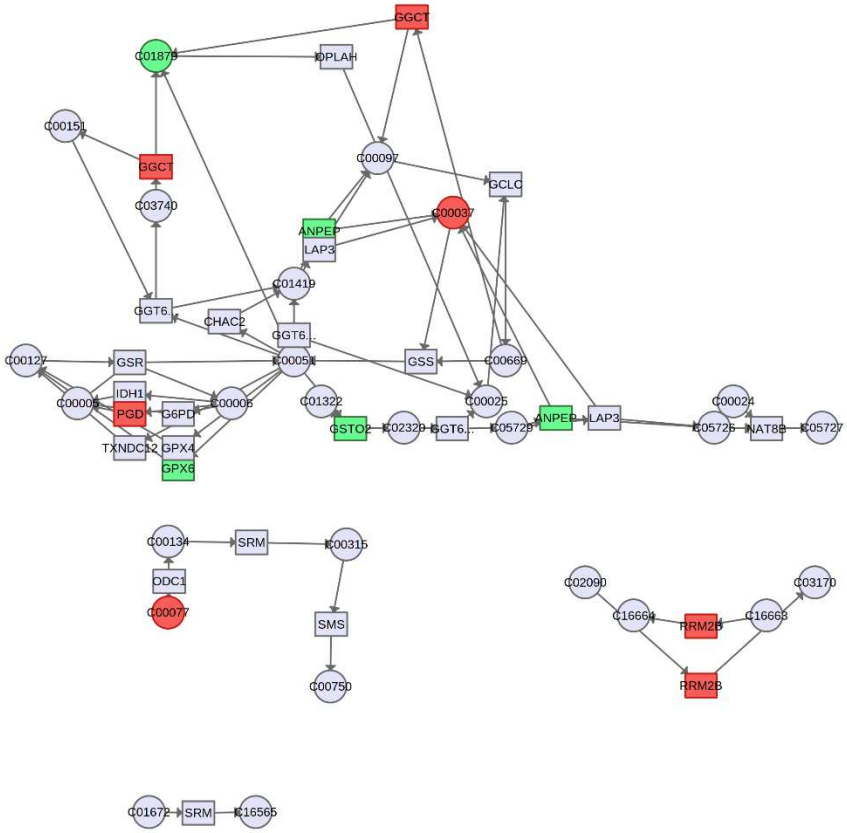

Figure 4.7: KEGG metabolic pathway co-enrichment analysis. Glutathione metabolism.
Red indicates higher expression level (genes) or concentration (metabolites) in the tumor tissue and green indicates lower levels.

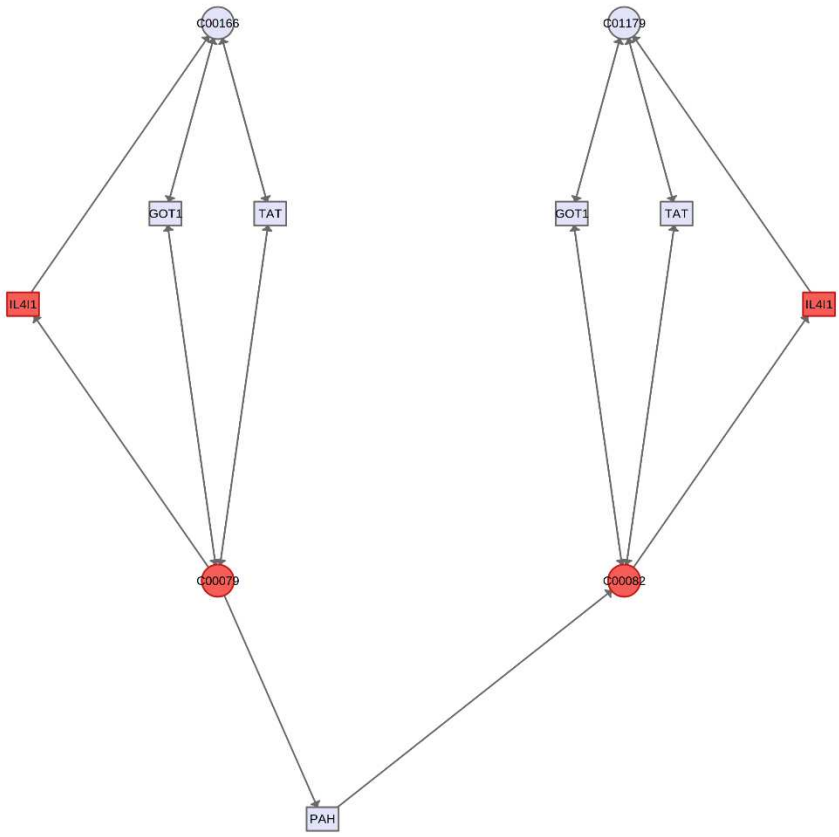

Figure 4.8: KEGG metabolic pathway co-enrichment analysis. Phenylalanine, tyrosine, and tryptophan biosynthesis pathway.
Red indicates higher expression level (genes) or concentration (metabolites) in the tumor tissue and green indicates lower levels.

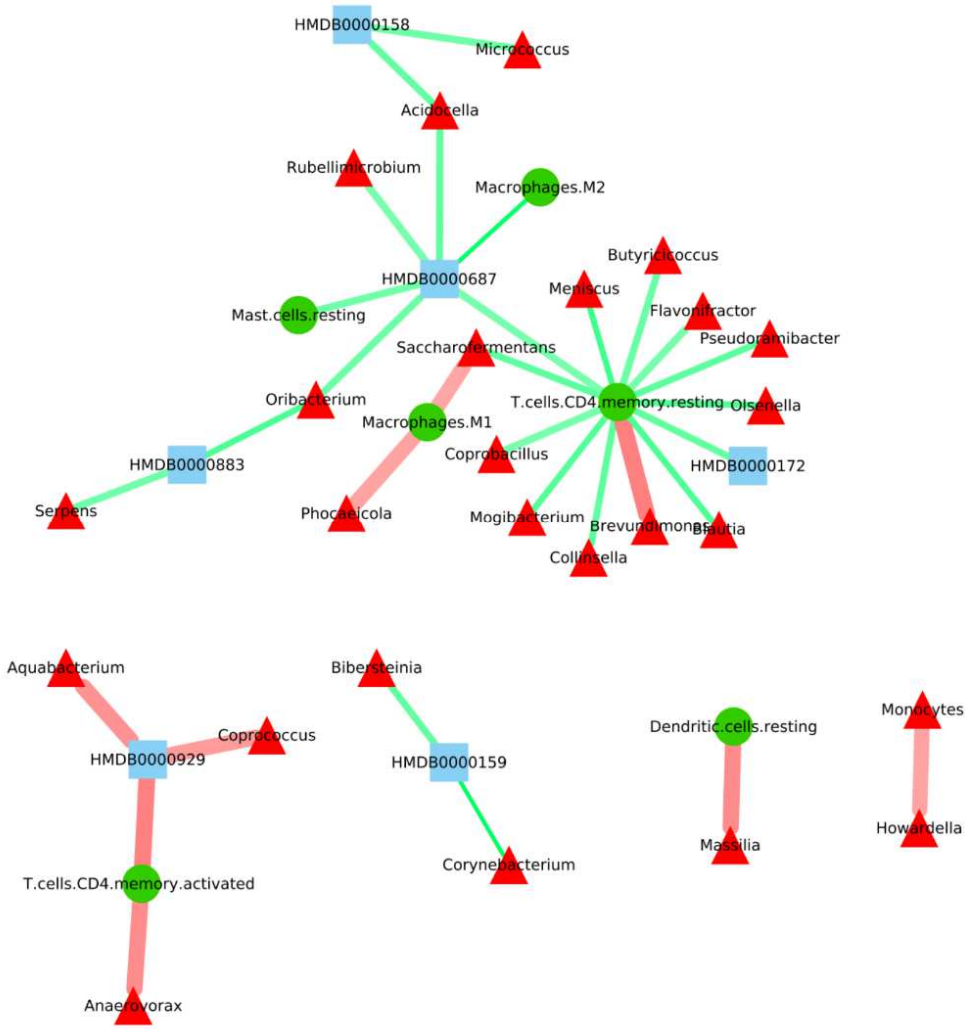

Figure 4.9: Network analysis.
Network analysis of essential amino acids, microbes, and immune cells. The blue square indicates metabolites, the red triangle indicates microbes, and the green circles indicate immune cells. Red edges indicate negative relationships and green edges indicate positive relationships, the thickness of the edges indicates the correlation coefficient, whereas thicker lines mean a higher correlation coefficient.

4.4 Discussion and Conclusion

We took a systems biology approach in this study and analyzed microbiota – metabolite – immune cell interactions in colorectal cancer. While previous studies have analyzed pair-wise interactions, limited studies have explored the 3 – way interactions. More importantly, we found a central role of amino acids in the interactions that may have implications in developing treatment or diagnostic strategies.

Previous studies have similarly found higher levels of essential amino acids in the CRC tumor tissues, fecal samples, and even blood samples[88,230–234]. Typically, food-source amino acids are absorbed in the small intestines. In certain situations, for example, high protein diet, undigested proteins can pass through the small intestines and end up in the large intestines, where proteolytic bacteria can break them down through the putrefaction process[231]. Studies have shown that this process can produce harmful metabolites that contribute to CRC progression. It is also possible that due to faster proliferation and higher levels of misfolded proteins being produced by the tumor cells, they require a larger amount of amino acids overall to the tumor location, leading to higher detected concentrations[235]. Another possible contributor to the amino acids in the gut microbiota, where many species are known to produce all 20 amino acids needed by humans[236].

Higher levels of these essential amino acids in the tumor tissue may have several consequences. While we have identified multiple relationships between

metabolites, microbes, and immune cell fractions, limited studies have shown the same relationships. One notable essential amino acid is tryptophan. Prior studies have found that IL4I1 in tumor potentially act as a novel immune checkpoint. This gene act through mediating tryptophan metabolism which influences M2 macrophages polarization and inhibits both T-cells and B-cells in the tumor microenvironment[237–239]. This is consistent with our analysis (**Figure 4.9**) showing that tryptophan is negatively correlated with activated CD4 T Cells. Additionally, our co-pathway analysis has also found enrichment of the Phenylalanine, tyrosine, and tryptophan biosynthesis pathway in the tumor tissues (**Figure 4.8**), where phenylalanine, tyrosine, and IL4I1 contribute to the pathway enrichment. Additionally, we found that tyrosine is positively correlated with *Micrococcus*, which has shown to be able to produce tyrosine[240]. Moreover, phenylalanine, which is found at a higher level in the tumor tissues and an essential amino acid that can be converted into tyrosine in humans, is negatively correlated with *Corynebacterium. Corynebacterium* has long been thought to be able to elicit an immune response and potentially useful in cancer treatment[241–244].

Taken together, evidences suggest a potential pathway where *Micrococcus* produces tyrosine in the tumor microenvironment, which through IL4I1 contributes to immune cell suppression and finally contributes to tumor progression [237–239]. We postulate that microbial-derived metabolites act as an intermediate in the tumor microenvironment that can promote tumor progression through both nutrient production and immune cell regulation. The contribution of

the gut microbiota in sustaining metabolite conditioning to promote immunosuppression is relatively novel and unexplored. Further investigations on the metabolic mediators secreted by the gut microbiota will undoubtedly provide a new level of understanding of immune homeostasis and novel methods in modulating immune responses in a multitude of diseases including improving the effectiveness of immune-checkpoint blockade therapies in cancers.

4.5 PUBLICATIONS

This chapter has been modified (with permission) from the published articles:

26. Yuan, C., Steer, C. J. & Subramanian, S. Host−microRNA−microbiota interactions in colorectal cancer. *Genes* (Basel) 10, (2019). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6523287/

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

We are still at the beginning of understanding many factors surrounding CRC, including microbiota and the immune system. These factors do not act alone, they form a complex network of interactions and together contribute to CRC development and progression[26]. It is thus important to develop and use combinatorial models rather than single-factor models, to examine the effect of multiple factors in preclinical testing. Although this may go against the traditional bottom-up approach, experience shows the necessity of using combinatorial models and technological advances support the development of such models.

This current dissertation outlined interactions between the host and microbiota, through the lens of miRNAs, metabolites, and immune cells, using a systems biology approach. We presented multiple interactions that warrant additional investigations. A recent effort to develop a mouse model with humanized microbiota may further propel the field of tumor-microbiota metabolic interactions (TMMI)[245]. We believe future research will require incorporating humanized microbiota animal models with a dual approach that leverages high-throughput genomics and metabolomics technologies is imperative in studying TMMI in CRC. Here we discuss a microbiota model that can be used to assess tumor growth and immune response.

Typically, to establish a humanized microbiota-associated (HMA) animal model, human microbiota samples are transferred into germ-free animals [168]. This method has shown short-term stability in retaining human microbiota in these animals. However, this method requires significant capital investment in

the dedicated germ-free facility and is not available to many researchers. Another method that simplifies the process uses extensive antibiotics treatment to condition the mouse colon, followed by a single dose of human microbiota gavage as depicted in **Fig. 5.1**[245]. This model is much more feasible compared to the germ-free model and has shown a similar ability to stably retain human microbiota in these animals over 21 days. More importantly, this method does not elicit systemic nor mucosal immune activation, suggesting potential long-term stability and relevance for evaluating immunotherapies. In light of the recent studies showing the importance of microbiota in enhancing immunotherapy response, the HMA model is needed for future immunotherapy testing. We are just beginning to scratch the surface of our understanding of the microbiome's contribution to health and disease but recent rapid advances that have yielded insight into therapeutic responses are promising that modulating the microbiome will provide a novel approach for patients with CRC. As a result of recent investigations that have implicated the microbiota contributing to the efficacy of immunotherapies, many researchers have sought to identify ways to manipulate the microbiome for improved responses to therapies. One method gaining noted attention comes from fecal microbiota transplantation (FMT) which takes patients' stool samples that had positive responses to immunotherapy and are transferred and retained in patients to assess the efficacy of immunotherapies. For more information on this, we direct readers to three excellent reviews on this topic[246–248].

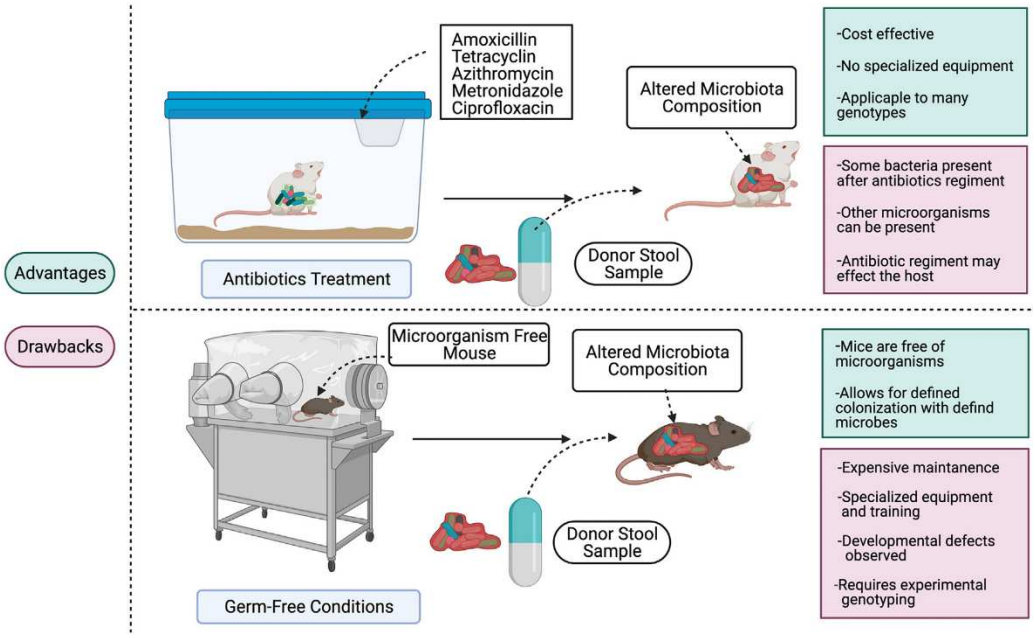

Figure 5.1: Methods of generating gnotobiotic mouse models that can be used to model human microbiota composition.
Advantages and drawbacks comparing antibiotic methods and germ-free conditions are included.

With a better understanding of how the microbiome is contributing to tumor progression and therapeutic outcomes, several clinical trials investigating FMT combination treatment with chemotherapy and immunotherapy are being explored[249]. While the majority of investigations have explored refractory melanoma, there are a number of trials looking at gastrointestinal cancers and specifically CRC. For example, NCT04130763, a phase 1 clinical trial sponsored by Peking University, is being conducted to examine combination treatments of FMT with anti-PD-L1 in patients with GI malignancies who were previously unresponsive to PD-L1 blockade. Further investigations examining the feasibility of Microbial Ecosystem Therapeutics (MET-4) such as NCT03686202. Unlike donor stool used in fecal transplants, which are incompletely characterized by complex communities of microbes and associated metabolites and fecal material, MET-4 consists of a defined mixture of pure live cultures of intestinal bacteria isolated from a stool sample of a healthy donor.

Other attempts at manipulating the microbiome for improved efficacy come in the form of supplemental probiotics. Several completed clinical trials have investigated probiotic treatment in the context of anti-tumor immunity in CRC[249]. Including NCT0372641, and NCT03782428 which investigated the role of probiotic functional foods in reducing CRC-related inflammatory markers and symptom alleviation. With these recent investigations showing that the microbiome can contribute to therapeutic response, it is clear that preclinical models are essential to recapitulate human microbiota in murine models to narrow down which bacteria are drivers in response and resistance. These

models require robust maintenance of a human microbial signature for relevant comparisons. Challenges that remain in assessing the role of the microbiota include the vast diversity of microbes in the gut as well as environmental and dietary differences amongst cohorts being studied. These factors can somewhat be controlled for in preclinical models but once scaled to clinical trials becomes a very complex issue that needs to be resolved.

Since Mice have been extensively used for preclinical studies and have provided invaluable results that have helped test many life-saving therapies for CRC patients, here we present a summary of the advantages and limitations of various surgical, genetic, and chemically induced mouse models for studying CRC (**Table 5.1**). At this moment, no single mouse model is better than another to study CRC. Depending on the focus of the study and the technical expertise of the researcher, one model may be preferred over another. Many models we discussed here require either extensive surgical expertise (orthotopic and humanized models) or expensive and dedicated equipment (orthotopic and human microbiota-associated models). These factors may also affect the choice of mouse models for a certain study. Besides mice, rats and pigs are also used for studying CRC [250,251]. Similarly, these models all have their advantages and limitations. It is important for a researcher to carefully balance the research focus and limitations of the animals.

Table 5.1: Summary of the advantages and limitations of different mouse models for CRC.

	Advantages	Limitations
Subcutaneous implantation	• Simple to establish • Easy to monitor	• Tumor not grown in the correct location and environment
Orthotopic implantation	• Tumor grown in the correct location and environment	• Requires extensive surgical training • Difficult to monitor
Organoids model	• Captures the benefits of both cell line and tumor chunks	• Expensive
Humanized model	• Human immune system	• Requires extensive surgical training • Expensive • Inconsistent humanization
Genetic model	• Tumor grown in the correct location and environment • Well established • Readily available	• Multiple genetic mutations are difficult to obtain • Can be very expensive • Difficult to monitor
Chemically induced model	• Tumor grown in the correct location and environment • Captures human CRC mutation profile	• Immunological features may not mimic human CRC • Chemicals are expensive • May alter the microbiota • Difficult to monitor
Human Microbiota Associated model	• Humanized microbiota	• The “healthy” microbiota composition is not defined • Long-term stability is unclear
Combinatorial models	• Capture advantages of multiple models	• May be difficult to establish • Maybe expensive • Many unknowns

In the future, alternatives to using animals in disease studies may become more feasible. Currently, organoid models represent a significant advance in modeling organs in vitro. It will not only reduce and replace the number of animals used in research, but it is also important in supporting the development of precision medicine. Researchers are also developing organ-on-chip models for more complex in vitro models as well as in silico models [252,253]. These animal-free models will undoubtedly have a significant impact on the drug discovery process.

5.1 PUBLICATIONS

This chapter has been modified (with permission) from the published articles[26,125,126]:

26. Yuan, C., Steer, C. J. & Subramanian, S. Host−microrna−microbiota interactions in colorectal cancer. *Genes* (Basel) 10, (2019).
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6523287/

125. Yuan, C. & Subramanian, S. MicroRNA Mediated Tumor-Microbiota Metabolic Interactions in Colorectal Cancer. *DNA Cell Biol* (2019).

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6477581/

126. Yuan, C. et al. Tumor models to assess immune response and tumor-microbiome interactions in colorectal cancer. *Pharmacol. Ther.* 231, 107981 (2022).

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8844062/

Bibliography

1. Turnbaugh, P. J. *et al.* The human microbiome project. *Nature* **449**, 804–810 (2007).
2. Burns, M. B. *et al.* Colorectal cancer mutational profiles correlate with defined microbial communities in the tumor microenvironment. *PLoS Genet.* **14**, e1007376 (2018).
3. Yuan, C., Burns, M. B., Subramanian, S. & Blekhman, R. Interaction between Host MicroRNAs and the Gut Microbiota in Colorectal Cancer. *mSystems* **3**, e00205-17 (2018).
4. Dejea, C. M. *et al.* Microbiota organization is a distinct feature of proximal colorectal cancers. *Proc Natl Acad Sci USA* **111**, 18321–18326 (2014).
5. Kostic, A. D. *et al.* Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res.* **22**, 292–298 (2012).
6. Shah, M. S. *et al.* Re-purposing 16S rRNA gene sequence data from within case paired tumor biopsy and tumor-adjacent biopsy or fecal samples to identify microbial markers for colorectal cancer. *PLoS ONE* **13**, e0207002 (2018).
7. García-Castillo, V., Sanhueza, E., McNerney, E., Onate, S. A. & García, A. Microbiota dysbiosis: a new piece in the understanding of the carcinogenesis puzzle. *J. Med. Microbiol.* **65**, 1347–1362 (2016).
8. Blekhman, R. *et al.* Host genetic variation impacts microbiome composition across human body sites. *Genome Biol.* **16**, 191 (2015).
9. Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* **159**, 789–799 (2014).
10. Bongers, G. *et al.* Interplay of host microbiota, genetic perturbations, and inflammation promotes local development of intestinal neoplasms in mice. *J. Exp. Med.* **211**, 457–472 (2014).
11. Davenport, E. R. *et al.* Genome-Wide Association Studies of the Human Gut Microbiota. *PLoS ONE* **10**, e0140301 (2015).
12. Goodrich, J. K. *et al.* Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe* **19**, 731–743 (2016).
13. Boleij, A. *et al.* The Bacteroides fragilis toxin gene is prevalent in the colon mucosa of colorectal cancer patients. *Clin. Infect. Dis.* **60**, 208–215 (2015).
14. Toprak, N. U. *et al.* A possible role of Bacteroides fragilis enterotoxin in the aetiology of colorectal cancer. *Clin. Microbiol. Infect.* **12**, 782–786 (2006).

15. Wu, S. *et al.* The Bacteroides fragilis toxin binds to a specific intestinal epithelial cell receptor. *Infect. Immun.* **74**, 5382–5390 (2006).

16. Kostic, A. D. *et al.* Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **14**, 207–215 (2013).

17. Mima, K. *et al.* Fusobacterium nucleatum in Colorectal Carcinoma Tissue According to Tumor Location. *Clin. Transl. Gastroenterol.* **7**, e200 (2016).

18. Mima, K. *et al.* Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. *Gut* **65**, 1973–1980 (2016).

19. Abed, J. *et al.* Fap2 Mediates Fusobacterium nucleatum Colorectal Adenocarcinoma Enrichment by Binding to Tumor-Expressed Gal-GalNAc. *Cell Host Microbe* **20**, 215–225 (2016).

20. Rubinstein, M. R. *et al.* Fusobacterium nucleatum promotes colorectal cancer by inducing Wnt/β-catenin modulator Annexin A1. *EMBO Rep.* **20**, (2019).

21. Rubinstein, M. R. *et al.* Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. *Cell Host Microbe* **14**, 195–206 (2013).

22. Schroeder, B. O. & Bäckhed, F. Signals from the gut microbiota to distant organs in physiology and disease. *Nat. Med.* **22**, 1079–1089 (2016).

23. Burns, M. B., Lynch, J., Starr, T. K., Knights, D. & Blekhman, R. Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment. *Genome Med.* **7**, 55 (2015).

24. Dove, W. F. *et al.* Intestinal neoplasia in the ApcMin mouse: independence from the microbial and natural killer (beige locus) status. *Cancer Res.* **57**, 812–814 (1997).

25. Zackular, J. P., Baxter, N. T., Chen, G. Y. & Schloss, P. D. Manipulation of the gut microbiota reveals role in colon tumorigenesis. *mSphere* **1**, (2016).

26. Yuan, C., Steer, C. J. & Subramanian, S. Host¯microrna¯microbiota interactions in colorectal cancer. *Genes (Basel)* **10**, (2019).

27. Wang, H.-F. *et al.* Evaluation of antibody level against Fusobacterium nucleatum in the serological diagnosis of colorectal cancer. *Sci. Rep.* **6**, 33440 (2016).

28. Louis, P., Hold, G. L. & Flint, H. J. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat. Rev. Microbiol.* **12**, 661–672 (2014).

29. Donohoe, D. R. *et al.* The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell Metab.* **13**, 517–526 (2011).

30. Vivarelli, S. *et al.* Gut microbiota and cancer: from pathogenesis to therapy. *Cancers (Basel)* **11**, (2019).

31. Blacher, E., Levy, M., Tatirovsky, E. & Elinav, E. Microbiome-Modulated Metabolites at the Interface of Host Immunity. *J. Immunol.* **198**, 572–580 (2017).

32. O'Keefe, S. J. D. Diet, microorganisms and their metabolites, and colon cancer. *Nat. Rev. Gastroenterol. Hepatol.* **13**, 691–706 (2016).

33. Johnson, C. H., Spilker, M. E., Goetz, L., Peterson, S. N. & Siuzdak, G. Metabolite and microbiome interplay in cancer immunotherapy. *Cancer Res.* **76**, 6146–6152 (2016).

34. Tjalsma, H., Boleij, A., Marchesi, J. R. & Dutilh, B. E. A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. *Nat. Rev. Microbiol.* **10**, 575–582 (2012).

35. Pate, K. T. *et al.* Wnt signaling directs a metabolic program of glycolysis and angiogenesis in colon cancer. *EMBO J.* **33**, 1454–1473 (2014).

36. Konsavage, W. M., Kyler, S. L., Rennoll, S. A., Jin, G. & Yochum, G. S. Wnt/β-catenin signaling regulates Yes-associated protein (YAP) gene expression in colorectal carcinoma cells. *J. Biol. Chem.* **287**, 11730–11739 (2012).

37. Strillacci, A. *et al.* Loss of miR-101 expression promotes Wnt/β-catenin signalling pathway activation and malignancy in colon cancer cells. *J. Pathol.* **229**, 379–389 (2013).

38. Bienz, M. & Clevers, H. Linking colorectal cancer to Wnt signaling. *Cell* **103**, 311–320 (2000).

39. Wu, S., Morin, P. J., Maouyo, D. & Sears, C. L. Bacteroides fragilis enterotoxin induces c-Myc expression and cellular proliferation. *Gastroenterology* **124**, 392–400 (2003).

40. Wu, S., Lim, K. C., Huang, J., Saidi, R. F. & Sears, C. L. Bacteroides fragilis enterotoxin cleaves the zonula adherens protein, E-cadherin. *Proc Natl Acad Sci USA* **95**, 14979–14984 (1998).

41. Diosdado, B. *et al.* MiR-17-92 cluster is associated with 13q gain and c-myc expression during colorectal adenoma to adenocarcinoma progression. *Br. J. Cancer* **101**, 707–714 (2009).

42. Mogilyansky, E. & Rigoutsos, I. The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death Differ.* **20**, 1603–1614 (2013).

43. O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V. & Mendell, J. T. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* **435**, 839–843 (2005).

44. Dang, C. V. c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Mol. Cell. Biol.* **19**, 1–11 (1999).

45. Yang, Y. *et al.* Fusobacterium nucleatum Increases Proliferation of Colorectal Cancer Cells and Tumor Development in Mice by Activating Toll-Like Receptor 4 Signaling to Nuclear Factor-κB, and Up-regulating Expression of MicroRNA-21. *Gastroenterology* **152**, 851-866.e24 (2017).

46. Dalmasso, G., Cougnoux, A., Delmas, J., Darfeuille-Michaud, A. & Bonnet, R. The bacterial genotoxin colibactin promotes colon tumor growth by modifying the tumor microenvironment. *Gut Microbes* **5**, 675–680 (2014).

47. Cougnoux, A. *et al.* Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype. *Gut* **63**, 1932–1942 (2014).

48. Baek, S. H. A novel link between SUMO modification and cancer metastasis. *Cell Cycle* **5**, 1492–1495 (2006).

49. Sicard, J.-F., Le Bihan, G., Vogeleer, P., Jacques, M. & Harel, J. Interactions of Intestinal Bacteria with Components of the Intestinal Mucus. *Front. Cell. Infect. Microbiol.* **7**, 387 (2017).

50. Li, S., Konstantinov, S. R., Smits, R. & Peppelenbosch, M. P. Bacterial biofilms in colorectal cancer initiation and progression. *Trends Mol. Med.* **23**, 18–30 (2017).

51. De Weirdt, R. & Van de Wiele, T. Micromanagement in the gut: microenvironmental factors govern colon mucosal biofilm structure and functionality. *npj Biofilms and Microbiomes* **1**, 15026 (2015).

52. Gordon, H. A. Morphological and physiological characterization of germfree life. *Ann. N. Y. Acad. Sci.* **78**, 208–220 (1959).

53. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**, 107–118 (2005).

54. Vétizou, M. *et al.* Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science* **350**, 1079–1084 (2015).

55. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).

56. Fabian, M. R. & Sonenberg, N. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat. Struct. Mol. Biol.* **19**, 586–593 (2012).

57. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).

58. Slattery, M. L. *et al.* Expression of Wnt-signaling pathway genes and their associations with miRNAs in colorectal cancer. *Oncotarget* **9**, 6075–6085 (2018).

59. Sarver, A. L. *et al.* Human colon cancer profiles show differential microRNA expression depending on mismatch repair status and are characteristic of undifferentiated proliferative states. *BMC Cancer* **9**, 401 (2009).

60. Sarver, A. L., Sarver, A. E., Yuan, C. & Subramanian, S. OMCD: oncomir cancer database. *BMC Cancer* **18**, 1223 (2018).

61. Wong, N. W., Chen, Y., Chen, S. & Wang, X. OncomiR: an online resource for exploring pan-cancer microRNA dysregulation. *Bioinformatics* **34**, 713–715 (2018).

62. Falzone, L. *et al.* Integrated analysis of colorectal cancer microRNA datasets: identification of microRNAs associated with tumor development. *Aging (Albany NY)* **10**, 1000–1014 (2018).

63. Li, Y. *et al.* Adenomatous polyposis coli (APC) regulates miR17-92 cluster through β-catenin pathway in colorectal cancer. *Oncogene* **35**, 4558–4568 (2016).

64. Koga, Y. *et al.* MicroRNA expression profiling of exfoliated colonocytes isolated from feces for colorectal cancer screening. *Cancer Prev Res (Phila Pa)* **3**, 1435–1442 (2010).

65. Fang, L. *et al.* MicroRNA-17-5p promotes chemotherapeutic drug resistance and tumour metastasis of colorectal cancer by repressing PTEN expression. *Oncotarget* **5**, 2974–2987 (2014).

66. Wu, C. W. *et al.* Detection of miR-92a and miR-21 in stool samples as potential screening biomarkers for colorectal cancer and polyps. *Gut* **61**, 739–745 (2012).

67. Kulda, V. *et al.* Relevance of miR-21 and miR-143 expression in tissue samples of colorectal carcinoma and its liver metastases. *Cancer Genet. Cytogenet.* **200**, 154–160 (2010).

68. Asangani, I. A. *et al.* MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene* **27**, 2128–2136 (2008).

69. Bullock, M. D. *et al.* Pleiotropic actions of miR-21 highlight the critical role of deregulated stromal microRNAs during colorectal cancer progression. *Cell Death Dis.* **4**, e684 (2013).

70. Li, L. *et al.* Sequential expression of miR-182 and miR-503 cooperatively targets FBXW7, contributing to the malignant transformation of colon adenoma to adenocarcinoma. *J. Pathol.* **234**, 488–501 (2014).

71. Dews, M. *et al.* The myc-miR-17~92 axis blunts TGF{beta} signaling and production of multiple TGF{beta}-dependent antiangiogenic factors. *Cancer Res.* **70**, 8233–8246 (2010).

72. Hu, S., Liu, L., Chang, E. B., Wang, J.-Y. & Raufman, J.-P. Butyrate inhibits pro-proliferative miR-92a by diminishing c-Myc-induced miR-17-92a cluster transcription in human colon cancer cells. *Mol. Cancer* **14**, 180 (2015).

73. Yamamichi, N. *et al.* Locked nucleic acid in situ hybridization analysis of miR-21 expression during colorectal cancer development. *Clin. Cancer Res.* **15**, 4009–4016 (2009).

74. Grillari, J., Hackl, M. & Grillari-Voglauer, R. miR-17-92 cluster: ups and downs in cancer and aging. *Biogerontology* **11**, 501–506 (2010).

75. Makinoshima, H. *et al.* Signaling through the Phosphatidylinositol 3-Kinase (PI3K)/Mammalian Target of Rapamycin (mTOR) Axis Is Responsible for Aerobic Glycolysis mediated by Glucose Transporter in Epidermal Growth Factor Receptor (EGFR)-mutated Lung Adenocarcinoma. *J. Biol. Chem.* **290**, 17495–17504 (2015).

76. Weijenberg, M. P. *et al.* The mTOR Pathway and the Role of Energy Balance Throughout Life in Colorectal Cancer Etiology and Prognosis: Unravelling Mechanisms Through a Multidimensional Molecular Epidemiologic Approach. *Curr. Nutr. Rep.* **2**, 19–26 (2013).

77. Kohlhapp, F. J., Mitra, A. K., Lengyel, E. & Peter, M. E. MicroRNAs as mediators and communicators between cancer cells and the tumor microenvironment. *Oncogene* **34**, 5857–5868 (2015).

78. Kosaka, N. *et al.* Trash or Treasure: extracellular microRNAs and cell-to-cell communication. *Front. Genet.* **4**, 173 (2013).

79. Runtsch, M. C., Round, J. L. & O'Connell, R. M. MicroRNAs and the regulation of intestinal homeostasis. *Front. Genet.* **5**, 347 (2014).

80. Nielsen, B. S. *et al.* High levels of microRNA-21 in the stroma of colorectal cancers predict short disease-free survival in stage II colon cancer patients. *Clin. Exp. Metastasis* **28**, 27–38 (2011).

81. Dews, M. *et al.* Augmentation of tumor angiogenesis by a Myc-activated microRNA cluster. *Nat. Genet.* **38**, 1060–1065 (2006).

82. Zhuang, G. *et al.* Tumour-secreted miR-9 promotes endothelial cell migration and angiogenesis by activating the JAK-STAT pathway. *EMBO J.* **31**, 3513–3523 (2012).

83. Fanini, F. & Fabbri, M. Cancer-derived exosomic microRNAs shape the immune system within the tumor microenvironment: State of the art. *Semin. Cell Dev. Biol.* **67**, 23–28 (2017).

84. Sonda, N. *et al.* miR-142-3p prevents macrophage differentiation during cancer-induced myelopoiesis. *Immunity* **38**, 1236–1249 (2013).

85. Zhang, M. *et al.* Both miR-17-5p and miR-20a alleviate suppressive potential of myeloid-derived suppressor cells by modulating STAT3 expression. *J. Immunol.* **186**, 4716–4724 (2011).

86. Mei, S. *et al.* MicroRNA-200c Promotes Suppressive Potential of Myeloid-Derived Suppressor Cells by Modulating PTEN and FOG2 Expression. *PLoS ONE* **10**, e0135867 (2015).

87. Warburg, O. On the origin of cancer cells. *Science* **123**, 309–314 (1956).

88. Weir, T. L. *et al.* Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. *PLoS ONE* **8**, e70803 (2013).

89. Hirayama, A. *et al.* Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. *Cancer Res.* **69**, 4918–4925 (2009).

90. Brown, D. G. *et al.* Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. *Cancer Metab.* **4**, 11 (2016).

91. Iwaya, T. *et al.* Downregulation of miR-144 is associated with colorectal cancer progression via activation of mTOR signaling pathway. *Carcinogenesis* **33**, 2391–2397 (2012).

92. Sun, Y., Zhao, X., Zhou, Y. & Hu, Y. miR-124, miR-137 and miR-340 regulate colorectal cancer growth via inhibition of the Warburg effect. *Oncol. Rep.* **28**, 1346–1352 (2012).

93. Taniguchi, K. *et al.* Positive feedback of DDX6/c-Myc/PTB1 regulated by miR-124 contributes to maintenance of the Warburg effect in colon cancer cells. *Biochim. Biophys. Acta* **1852**, 1971–1980 (2015).

94. Taniguchi, K. *et al.* MicroRNA-124 inhibits cancer cell growth through PTB1/PKM1/PKM2 feedback cascade in colorectal cancer. *Cancer Lett.* **363**, 17–27 (2015).

95. Wang, J. *et al.* Lactate dehydrogenase A negatively regulated by miRNAs promotes aerobic glycolysis and is increased in colorectal cancer. *Oncotarget* **6**, 19456–19468 (2015).

96. Hirschhaeuser, F., Sattler, U. G. A. & Mueller-Klieser, W. Lactate: a metabolic key player in cancer. *Cancer Res.* **71**, 6921–6925 (2011).

97. Singh, R., Yadav, V., Kumar, S. & Saini, N. MicroRNA-195 inhibits proliferation, invasion and metastasis in breast cancer cells by targeting FASN, HMGCR, ACACA and CYP27B1. *Sci. Rep.* **5**, 17454 (2015).

98. Mao, J. H. *et al.* microRNA-195 suppresses osteosarcoma cell invasion and migration in vitro by targeting FASN. *Oncol. Lett.* **4**, 1125–1129 (2012).

99. Long, X. H. *et al.* Tumor suppressive microRNA-424 inhibits osteosarcoma cell migration and invasion via targeting fatty acid synthase. *Experimental and Therapeutic Medicine* **5**, 1048–1052 (2013).

100. Buck, M. D., O'Sullivan, D. & Pearce, E. L. T cell metabolism drives immunity. *J. Exp. Med.* **212**, 1345–1360 (2015).

101. Wei, Z., Cui, L., Mei, Z., Liu, M. & Zhang, D. miR-181a mediates metabolic shift in colon cancer cells via the PTEN/AKT pathway. *FEBS Lett.* **588**, 1773–1779 (2014).

102. Schee, K. *et al.* Deep sequencing the microrna transcriptome in colorectal cancer. *PLoS ONE* **8**, e66165 (2013).

103. Song, B. *et al.* miR-192 Regulates dihydrofolate reductase and cellular proliferation through the p53-microRNA circuit. *Clin. Cancer Res.* **14**, 8080–8086 (2008).

104. Wang, Y., Tang, Q., Li, M., Jiang, S. & Wang, X. MicroRNA-375 inhibits colorectal cancer growth by targeting PIK3CA. *Biochem. Biophys. Res. Commun.* **444**, 199–204 (2014).

105. Dalmasso, G. *et al.* Microbiota modulate host gene expression via microRNAs. *PLoS ONE* **6**, e19293 (2011).

106. Peck, B. C. E. *et al.* Functional transcriptomics in diverse intestinal epithelial cell types reveals robust microrna sensitivity in intestinal stem cells to microbial status. *J. Biol. Chem.* **292**, 2586–2600 (2017).

107. Phua, L. C. *et al.* Global fecal microRNA profiling in the identification of biomarkers for colorectal cancer screening among Asians. *Oncol. Rep.* **32**, 97–104 (2014).

108. Rotelli, M. T. *et al.* Fecal microRNA profile in patients with colorectal carcinoma before and after curative surgery. *Int. J. Colorectal Dis.* **30**, 891–898 (2015).

109. Liu, S. *et al.* The host shapes the gut microbiota via fecal microrna. *Cell Host Microbe* **19**, 32–43 (2016).

110. Teng, Y. *et al.* Plant-Derived Exosomal MicroRNAs Shape the Gut Microbiota. *Cell Host Microbe* **24**, 637-652.e8 (2018).

111. Baier, S. R., Nguyen, C., Xie, F., Wood, J. R. & Zempleni, J. MicroRNAs are absorbed in biologically meaningful amounts from nutritionally relevant doses of cow milk and affect gene expression in peripheral blood mononuclear cells, HEK-293 kidney cell cultures, and mouse livers. *J. Nutr.* **144**, 1495–1500 (2014).

112. Izumi, H. *et al.* Bovine milk exosomes contain microRNA and mRNA and are taken up by human macrophages. *J. Dairy Sci.* **98**, 2920–2933 (2015).

113. Zhang, L. *et al.* Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. *Cell Res.* **22**, 107–126 (2012).

114. Mu, J. *et al.* Interspecies communication between plant and mouse gut host cells through edible plant derived exosome-like nanoparticles. *Mol. Nutr. Food Res.* **58**, 1561–1573 (2014).

115. Wolf, T., Baier, S. R. & Zempleni, J. The Intestinal Transport of Bovine Milk Exosomes Is Mediated by Endocytosis in Human Colon Carcinoma Caco-2 Cells and Rat Small Intestinal IEC-6 Cells. *J. Nutr.* **145**, 2201–2206 (2015).

116. Gopalakrishnan, V. *et al.* Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* **359**, 97–103 (2018).

117. Matson, V. *et al.* The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science* **359**, 104–108 (2018).

118. Sivan, A. *et al.* Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science* **350**, 1084–1089 (2015).

119. Gopalakrishnan, V., Helmink, B. A., Spencer, C. N., Reuben, A. & Wargo, J. A. The influence of the gut microbiome on cancer, immunity, and cancer immunotherapy. *Cancer Cell* **33**, 570–580 (2018).

120. Zelante, T. *et al.* Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. *Immunity* **39**, 372–385 (2013).

121. Gottesman, S. & Storz, G. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).

122. Lässer, C. *et al.* Human saliva, plasma and breast milk exosomes contain RNA: uptake by macrophages. *J. Transl. Med.* **9**, 9 (2011).

123. Deng, Z. *et al.* Broccoli-Derived Nanoparticle Inhibits Mouse Colitis by Activating Dendritic Cell AMP-Activated Protein Kinase. *Mol. Ther.* **25**, 1641–1654 (2017).

124. Ju, S. *et al.* Grape exosome-like nanoparticles induce intestinal stem cells and protect mice from DSS-induced colitis. *Mol. Ther.* **21**, 1345–1357 (2013).

125. Yuan, C. & Subramanian, S. MicroRNA Mediated Tumor-Microbiota Metabolic Interactions in Colorectal Cancer. *DNA Cell Biol* (2019).

126. Yuan, C. *et al.* Tumor models to assess immune response and tumor-microbiome interactions in colorectal cancer. *Pharmacol. Ther.* **231**, 107981 (2022).

127. Blaut, M. & Clavel, T. Metabolic diversity of the intestinal microbiota: implications for health and disease. *J. Nutr.* **137**, 751S–5S (2007).

128. Abdollahi-Roodsaz, S., Abramson, S. B. & Scher, J. U. The metabolic role of the gut microbiota in health and rheumatic disease: mechanisms and interventions. *Nat. Rev. Rheumatol.* **12**, 446–455 (2016).

129. Belkaid, Y. & Hand, T. W. Role of the microbiota in immunity and inflammation. *Cell* **157**, 121–141 (2014).

130. Vangay, P., Ward, T., Gerber, J. S. & Knights, D. Antibiotics, pediatric dysbiosis, and disease. *Cell Host Microbe* **17**, 553–564 (2015).

131. Nicholson, J. K. *et al.* Host-gut microbiota metabolic interactions. *Science* **336**, 1262–1267 (2012).

132. Said, H. M. & Mohammed, Z. M. Intestinal absorption of water-soluble vitamins: an update. *Curr. Opin. Gastroenterol.* **22**, 140–146 (2006).

133. Scheppach, W. Effects of short chain fatty acids on gut morphology and function. *Gut* **35**, S35-8 (1994).

134. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).

135. Samuel, B. S. *et al.* Genomic and metabolic adaptations of Methanobrevibacter smithii to the human gut. *Proc Natl Acad Sci USA* **104**, 10643–10648 (2007).

136. Koeth, R. A. *et al.* Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.* **19**, 576–585 (2013).

137. Gu, S. *et al.* Bacterial community mapping of the mouse gastrointestinal tract. *PLoS ONE* **8**, e74957 (2013).

138. Stanley, D., Hughes, R. J. & Moore, R. J. Microbiota of the chicken gastrointestinal tract: influence on health, productivity and disease. *Appl. Microbiol. Biotechnol.* **98**, 4301–4310 (2014).

139. Gong, J. *et al.* 16S rRNA gene-based analysis of mucosa-associated bacterial community and phylogeny in the chicken gastrointestinal tracts: from crops to ceca. *FEMS Microbiol. Ecol.* **59**, 147–157 (2007).

140. Suchodolski, J. S., Camacho, J. & Steiner, J. M. Analysis of bacterial diversity in the canine duodenum, jejunum, ileum, and colon by comparative 16S rRNA gene analysis. *FEMS Microbiol. Ecol.* **66**, 567–578 (2008).

141. Mao, S., Zhang, M., Liu, J. & Zhu, W. Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: membership and potential function. *Sci. Rep.* **5**, 16116 (2015).

142. Stearns, J. C. *et al.* Bacterial biogeography of the human digestive tract. *Sci. Rep.* **1**, 170 (2011).

143. Jeraldo, P. *et al.* IM-TORNADO: a tool for comparison of 16S reads from paired-end libraries. *PLoS ONE* **9**, e114804 (2014).

144. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).

145. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).

146. Lozupone, C. A., Hamady, M., Kelley, S. T. & Knight, R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl. Environ. Microbiol.* **73**, 1576–1585 (2007).

147. Bray, J. R. & Curtis, J. T. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957).

148. Clarke, K. R. Non-parametric multivariate analyses of changes in community structure. *Austral Ecol.* **18**, 117–143 (1993).

149. McHardy, I. H. *et al.* Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. *Microbiome* **1**, 17 (2013).

150. Wang, Y. *et al.* Topographical variation in metabolic signatures of human gastrointestinal biopsies revealed by high-resolution magic-angle spinning 1H NMR spectroscopy. *J. Proteome Res.* **6**, 3944–3951 (2007).

151. Claus, S. P. *et al.* Systemic multicompartmental effects of the gut microbiome on mouse metabolic phenotypes. *Mol. Syst. Biol.* **4**, 219 (2008).

152. Kong, F. *et al.* Gut microbiota signatures of longevity. *Curr. Biol.* **26**, R832–R833 (2016).

153. Gorbach, S. L. Probiotics and gastrointestinal health. *Am. J. Gastroenterol.* **95**, S2-4 (2000).

154. Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638 (2005).

155. McKenna, P. *et al.* The macaque gut microbiome in health, lentiviral infection, and chronic enterocolitis. *PLoS Pathog.* **4**, e20 (2008).

156. Harrell, L. *et al.* Standard colonic lavage alters the natural state of mucosal-associated microbiota in the human colon. *PLoS ONE* **7**, e32545 (2012).

157. Blekhman, R. *et al.* Comparative metabolomics in primates reveals the effects of diet and gene regulatory variation on metabolic divergence. *Sci. Rep.* **4**, 5809 (2014).

158. Martin, F.-P. J. *et al.* Topographical variation in murine intestinal metabolic profiles in relation to microbiome speciation and functional ecological activity. *J. Proteome Res.* **8**, 3464–3474 (2009).

159. Kwak, M.-K. & Kensler, T. W. Targeting NRF2 signaling for cancer chemoprevention. *Toxicol. Appl. Pharmacol.* **244**, 66–76 (2010).

160. Yuan, C., Graham, M., Staley, C. & Subramanian, S. Mucosal Microbiota and Metabolome along the Intestinal Tract Reveal a Location-Specific Relationship. *mSystems* **5**, (2020).

161. Candon, S. *et al.* Antibiotics in early life alter the gut microbiome and increase disease incidence in a spontaneous mouse model of autoimmune insulin-dependent diabetes. *PLoS ONE* **10**, e0125448 (2015).

162. Tremaroli, V. & Bäckhed, F. Functional Interactions Between the Gut Microbiota and Host Metabolism. *Nature* **489**, 242–249 (2012).

163. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).

164. Faith, J. J. *et al.* The long-term stability of the human gut microbiota. *Science* **341**, 1237439 (2013).

165. Nakatsu, G. *et al.* Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat. Commun.* **6**, 8727 (2015).

166. Wang, T. *et al.* Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J.* **6**, 320–329 (2012).

167. Shen, X. J. *et al.* Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas. *Gut Microbes* **1**, 138–147 (2010).

168. Turnbaugh, P. J. *et al.* The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* **1**, 6ra14 (2009).

169. Ridaura, V. K. *et al.* Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**, 1241214 (2013).

170. Kesse, E., Clavel-Chapelon, F. & Boutron-Ruault, M. C. Dietary patterns and risk of colorectal tumors: a cohort of French women of the National Education System (E3N). *Am. J. Epidemiol.* **164**, 1085–1093 (2006).

171. Huxley, R. R. *et al.* The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence. *Int. J. Cancer* **125**, 171–180 (2009).

172. Kune, S., Kune, G. A. & Watson, L. F. Case□control study of dietary etiological factors: The Melbourne colorectal cancer study. *Nutrition and cancer* (1987).

173. Schulz, M. D. *et al.* High-fat-diet-mediated dysbiosis promotes intestinal carcinogenesis independently of obesity. *Nature* **514**, 508–512 (2014).

174. Goodrich, J. K., Davenport, E. R., Waters, J. L., Clark, A. G. & Ley, R. E. Cross-species comparisons of host genetic associations with the microbiome. *Science* **352**, 532–535 (2016).

175. Knights, D. *et al.* Complex host genetics influence the microbiome in inflammatory bowel disease. *Genome Med.* **6**, 107 (2014).

176. Uronis, J. M. *et al.* Modulation of the intestinal microbiota alters colitis-associated colorectal cancer susceptibility. *PLoS ONE* **4**, e6026 (2009).

177. Cai, L., Ye, L., Tong, A. H. Y., Lok, S. & Zhang, T. Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. *PLoS ONE* **8**, e53649 (2013).

178. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).

179. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).

180. Navas-Molina, J. A. *et al.* Advancing our understanding of the human microbiome using QIIME. *Meth. Enzymol.* **531**, 371–444 (2013).

181. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

182. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

183. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**, 31 (2012).

184. Anders, S., Pyl, P. T. & Huber, W. HTSeq — a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

185. Metpally, R. P. R. *et al.* Comparison of Analysis Tools for miRNA High Throughput Sequencing Using Nerve Crush as a Model. *Front. Genet.* **4**, 20 (2013).

186. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94 (2010).

187. Lopez, J. P. *et al.* Biomarker discovery: quantification of microRNAs and other small non-coding RNAs using next generation sequencing. *BMC Med. Genomics* **8**, 35 (2015).

188. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

189. Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8**, e1002687 (2012).

190. Langille, M. G. I. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**, 814–821 (2013).

191. Vlachos, I. S. *et al.* DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res.* **43**, W460-6 (2015).

192. Oberg, A. L. *et al.* miRNA expression in colon polyps provides evidence for a multihit model of colon cancer. *PLoS ONE* **6**, e20465 (2011).

193. Geng, J., Fan, H., Tang, X., Zhai, H. & Zhang, Z. Diversified pattern of the human colorectal cancer microbiome. *Gut Pathog.* **5**, 2 (2013).

194. Vogtmann, E. *et al.* Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing. *PLoS ONE* **11**, e0155362 (2016).

195. Yu, J. *et al.* Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* **66**, 70–78 (2017).

196. Dharmani, P., Strauss, J., Ambrose, C., Allen-Vercoe, E. & Chadee, K. Fusobacterium nucleatum infection of colonic cells stimulates MUC2 mucin and tumor necrosis factor alpha. *Infect. Immun.* **79**, 2597–2607 (2011).

197. Andrian, E., Grenier, D. & Rouabhia, M. Porphyromonas gingivalis gingipains mediate the shedding of syndecan-1 from the surface of gingival epithelial cells. *Oral Microbiol. Immunol.* **21**, 123–128 (2006).

198. Benvenuti, S. *et al.* Oncogenic activation of the RAS/RAF signaling pathway impairs the response of metastatic colorectal cancers to anti-epidermal growth factor receptor antibody therapies. *Cancer Res.* **67**, 2643–2648 (2007).

199. Hynes, N. E. & Lane, H. A. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat. Rev. Cancer* **5**, 341–354 (2005).

200. De Luca, A., Maiello, M. R., D'Alessio, A., Pergameno, M. & Normanno, N. The RAS/RAF/MEK/ERK and the PI3K/AKT signalling pathways: role in cancer pathogenesis and implications for therapeutic approaches. *Expert Opin. Ther. Targets* **16 Suppl 2**, S17-27 (2012).

201. Tahara, T. *et al.* Fusobacterium in colonic flora and molecular features of colorectal carcinoma. *Cancer Res.* **74**, 1311–1318 (2014).

202. Castellarin, M. *et al.* Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res.* **22**, 299–306 (2012).

203. Kholodkova, E. V., IuM, K. & Baturo, A. P. Etiologic role of bacteria of the genus Providencia in acute intestinal diseases. *Zhurnal Mikrobiologii …* (1977).

204. Shima, A., Hinenoya, A., Asakura, M., Nagita, A. & Yamasaki, S. Prevalence of Providencia strains among children with diarrhea in Japan. *Jpn J Infect Dis* **65**, 545–547 (2012).

205. Murata, T. *et al.* A large outbreak of foodborne infection attributed to Providencia alcalifaciens. *J. Infect. Dis.* **184**, 1050–1055 (2001).

206. Wu, S., Rhee, K.-J., Zhang, M., Franco, A. & Sears, C. L. Bacteroides fragilis toxin stimulates intestinal epithelial cell shedding and gamma-secretase-dependent E-cadherin cleavage. *J. Cell Sci.* **120**, 1944–1952 (2007).

207. Nougayrède, J.-P. *et al.* Escherichia coli induces DNA double-strand breaks in eukaryotic cells. *Science* **313**, 848–851 (2006).

208. Toller, I. M. *et al.* Carcinogenic bacterial pathogen Helicobacter pylori triggers DNA double-strand breaks and a DNA damage response in its host cells. *Proc Natl Acad Sci USA* **108**, 14944–14949 (2011).

209. He, T. C. *et al.* Identification of c-MYC as a target of the APC pathway. *Science* **281**, 1509–1512 (1998).

210. Mann, B. *et al.* Target genes of β-catenin–T cell-factor/lymphoid-enhancer-factor signaling in human colorectal carcinomas. *Proc. Natl. Acad. Sci. USA* **96**, 1603–1608 (1999).

211. Zhang, X., Gaspard, J. P. & Chung, D. C. Regulation of vascular endothelial growth factor by the Wnt and K-ras pathways in colonic neoplasia. *Cancer Res.* **61**, 6050–6054 (2001).

212. Linden, R. *et al.* Physiology of the prion protein. *Physiol. Rev.* **88**, 673–728 (2008).

213. Fang, J. Y. & Richardson, B. C. The MAPK signalling pathways and colorectal cancer. *Lancet Oncol.* **6**, 322–327 (2005).

214. Yamaguchi, H., Bhalla, K. & Wang, H.-G. Bax plays a pivotal role in thapsigargin-induced apoptosis of human colon cancer HCT116 cells by controlling Smac/Diablo and Omi/HtrA2 release from mitochondria. *Cancer Res.* **63**, 1483–1489 (2003).

215. Wynendaele, E. *et al.* Crosstalk between the microbiome and cancer cells by quorum sensing peptides. *Peptides* **64**, 40–48 (2015).

216. Tailford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. Mucin glycan foraging in the human gut microbiome. *Front. Genet.* **6**, 81 (2015).

217. Luca, F., Kupfer, S. S., Knights, D., Khoruts, A. & Blekhman, R. Functional Genomics of Host-Microbiome Interactions in Humans. *Trends Genet.* **34**, 30–40 (2018).

218. Leamy, L. J. *et al.* Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut

microbiome in an advanced intercross population of mice. *Genome Biol.* **15**, 552 (2014).

219. Davison, J. M. *et al.* Microbiota regulate intestinal epithelial gene expression by suppressing the transcription factor Hepatocyte nuclear factor 4 alpha. *Genome Res.* **27**, 1195–1206 (2017).

220. Camp, J. G. *et al.* Microbiota modulate transcription in the intestinal epithelium without remodeling the accessible chromatin landscape. *Genome Res.* **24**, 1504–1516 (2014).

221. Richards, A. L. *et al.* Genetic and transcriptional analysis of human host response to healthy gut microbiota. *mSystems* **1**, (2016).

222. Richards, A. L. *et al.* Gut microbiota has a widespread and modifiable effect on host gene regulation. *mSystems* **4**, (2019).

223. Shen, W. & Durum, S. K. Synergy of IL-23 and Th17 cytokines: new light on inflammatory bowel disease. *Neurochem. Res.* **35**, 940–946 (2010).

224. Zhao, X., Yuan, C., Wangmo, D. & Subramanian, S. Tumor-Secreted Extracellular Vesicles Regulate T-Cell Costimulation and Can Be Manipulated To Induce Tumor-Specific T-Cell Responses. *Gastroenterology* **161**, 560-574.e11 (2021).

225. Katoku-Kikyo, N. *et al.* Per1/Per2-Igf2 axis-mediated circadian regulation of myogenic differentiation. *J. Cell Biol.* **220**, (2021).

226. Sheng, Q.-S. *et al.* Comparison of gut microbiome in human colorectal cancer in paired tumor and adjacent normal tissues. *Onco Targets Ther* **13**, 635–646 (2020).

227. Gao, R. *et al.* Mucosa-associated microbiota signature in colorectal cancer. *Eur. J. Clin. Microbiol. Infect. Dis.* **36**, 2073–2083 (2017).

228. Flemer, B. *et al.* Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut* **66**, 633–643 (2017).

229. Chen, W., Liu, F., Ling, Z., Tong, X. & Xiang, C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS ONE* **7**, e39743 (2012).

230. Gao, P., Zhou, C., Zhao, L., Zhang, G. & Zhang, Y. Tissue amino acid profile could be used to differentiate advanced adenoma from colorectal cancer. *J. Pharm. Biomed. Anal.* **118**, 349–355 (2016).

231. Kaur, H., Das, C. & Mande, S. S. In silico analysis of putrefaction pathways in bacteria and its implication in colorectal cancer. *Front. Microbiol.* **8**, 2166 (2017).

232. Shen, Y. *et al.* Tissue metabolic profiling reveals major metabolic alteration in colorectal cancer. *Mol. Omics* **17**, 464–471 (2021).

233. Sirniö, P. *et al.* Alterations in serum amino-acid profile in the progression of colorectal cancer: associations with systemic inflammation, tumour stage and patient survival. *Br. J. Cancer* **120**, 238–246 (2019).

234. Lin, Y. *et al.* NMR-based fecal metabolomics fingerprinting as predictors of earlier diagnosis in patients with colorectal cancer. *Oncotarget* **7**, 29454–29464 (2016).

235. Medina, M. A., Márquez, J. & Núñez de Castro, I. Interchange of amino acids between tumor and host. *Biochem. Med. Metab. Biol.* **48**, 1–7 (1992).

236. Dai, Z.-L., Wu, G. & Zhu, W.-Y. Amino acid metabolism in intestinal bacteria: links between gut ecology and host health. *Front. Biosci.* **16**, 1768 (2011).

237. Yue, Y. *et al.* IL4I1 Is a Novel Regulator of M2 Macrophage Polarization That Can Inhibit T Cell Activation via L-Tryptophan and Arginine Depletion and IL-10 Production. *PLoS ONE* **10**, e0142979 (2015).

238. Sadik, A. *et al.* IL4I1 Is a Metabolic Immune Checkpoint that Activates the AHR and Promotes Tumor Progression. *Cell* **182**, 1252-1270.e34 (2020).

239. Molinier-Frenkel, V., Prévost-Blondel, A. & Castellano, F. The IL4I1 enzyme: A new player in the immunosuppressive tumor microenvironment. *Cells* **8**, (2019).

240. Voigt, M. N. & Eitenmiller, R. R. Production of Tyrosine and Histidine Decarboxylase by Dairy-Related Bacteria. *J. Food Prot.* **40**, 241–245 (1977).

241. Oettgen, H. F., Pinsky, C. M. & Delmonte, L. Treatment of Cancer with Immunomodulators: Corynebacterium parvum and Levamisole. *Medical Clinics of North America* **60**, 511–537 (1976).

242. Knapp, R. C. & Berkowitz, R. S. Corynebacterium parvum as an immunotherapeutic agent in an ovarian cancer model. *Am. J. Obstet. Gynecol.* **128**, 782–786 (1977).

243. Mattos-Guaraldi, A. L. *et al.* Corynebacterium diphtheriae threats in cancer patients. *Rev. Argent. Microbiol.* **33**, 96–100 (2001).

244. Marsh, C. L., Torrey, R. R., Woolley, J. L., Barker, G. R. & Lau, B. H. Superiority of intravesical immunotherapy with Corynebacterium parvum and Allium sativum in control of murine bladder cancer. *J. Urol.* **137**, 359–362 (1987).

245. Staley, C. *et al.* Stable engraftment of human microbiota into mice with a single oral gavage following antibiotic conditioning. *Microbiome* **5**, 87 (2017).

246. Matson, V., Chervin, C. S. & Gajewski, T. F. Cancer and the Microbiome-Influence of the Commensal Microbiota on Cancer, Immune Responses, and Immunotherapy. *Gastroenterology* **160**, 600–613 (2021).

247. Chen, D., Wu, J., Jin, D., Wang, B. & Cao, H. Fecal microbiota transplantation in cancer management: Current status and perspectives. *Int. J. Cancer* **145**, 2021–2031 (2019).

248. Zhou, C.-B., Zhou, Y.-L. & Fang, J.-Y. Gut microbiota in cancer immune response and immunotherapy. *Trends Cancer* (2021) doi:10.1016/j.trecan.2021.01.010.

249. Soto Chervin, C. & Gajewski, T. F. Microbiome-based interventions: therapeutic strategies in cancer immunotherapy. *Immuno-Oncology Technology* **8**, 12–20 (2020).

250. Schachtschneider, K. M. *et al.* The oncopig cancer model: an innovative large animal translational oncology platform. *Front. Oncol.* **7**, 190 (2017).

251. Sacksteder, M. R. Occurrence of spontaneous tumors in the germfree F344 rat. *J Natl Cancer Inst* **57**, 1371–1373 (1976).

252. Ahmad, A. A. *et al.* Optimization of 3-D organotypic primary colonic cultures for organ-on-chip applications. *J. Biol. Eng.* **8**, 9 (2014).

253. Terstappen, G. C. & Reggiani, A. In silico research in drug discovery. *Trends Pharmacol. Sci.* **22**, 23–26 (2001).

Appendix

According to the policy of *Genes* (https://www.mdpi.com/openaccess), I do not need to obtain permission for reusing the article titled *Host−microrna−microbiota interactions in colorectal cancer.,* as a part of my dissertation.

According to the policy of *DNA Cell Biology* (https://home.liebertpub.com/authors/permissions-and-reprints/197), I do not need to obtain permission for reusing the article titled *MicroRNA Mediated Tumor-Microbiota Metabolic Interactions in Colorectal Cancer*., as a part of my dissertation.

According to the policy of Pharmacology & Therapeutics (https://www.elsevier.com/about/policies/copyright/permissions), I do not need to obtain permission for reusing the article titled *Tumor models to assess immune response and tumor-microbiome interactions in colorectal cancer*., as a part of my dissertation.

According to the policy of *mSystems* (https://journals.asm.org/licenses), I do not need to obtain permission for reusing the article titled *Mucosal Microbiota and Metabolome along the Intestinal Tract Reveal a Location-Specific Relationship*., as a part of my dissertation.

According to the policy of *mSystems* (https://journals.asm.org/licenses), I do not need to obtain permission for reusing the article titled *Interaction between Host MicroRNAs and the Gut Microbiota in Colorectal Cancer*., as a part of my dissertation.

Table AT1: Unfiltered OTU table used for analysis. Accessed through: https://journals.asm.org/doi/suppl/10.1128/mSystems.00055-20/suppl_file/msystems.00055-20-st003.txt

Table AT2: Unfiltered untargeted metabolomic data used for analysis. Accessed through: https://journals.asm.org/doi/suppl/10.1128/mSystems.00055-20/suppl_file/msystems.00055-20-st004.xlsx

Table AT3: Baboon metadata, differentially abundant metabolites in the small and large intestine, and Ingenuity pathway analysis. Accessed through: https://journals.asm.org/doi/suppl/10.1128/mSystems.00055-20/suppl_file/msystems.00055-20-st001.xlsx

Table AT4: Spearman's correlation results for microbiota-metabolome interactions in the small and large intestine. Accessed through: https://journals.asm.org/doi/suppl/10.1128/mSystems.00055-20/suppl_file/msystems.00055-20-st002.xlsx

Table AT5: Full metadata associated with patients and a list of differentially expressed miRNAs. The tab "Full metadata" contains full patient clinical information, including age, gender, tumor location, tumor stage, and microsatellite stability (MSS) status, as well as sequencing quality information. The tab "Differentially expressed miRNAs" contains a full list of differentially expressed miRNAs identified using DESeq2 and for each miRNA includes the identification, log fold change, and q value and whether it is more highly expressed in the tumor or normal samples. Accessed through: https://journals.asm.org/doi/suppl/10.1128/mSystems.00205-17/suppl_file/sys003182230st1.xlsx

Table AT6: Unfiltered OTU table with relative abundances of the taxa used in the analysis. Accessed through: https://journals.asm.org/doi/suppl/10.1128/mSystems.00205-17/suppl_file/sys003182230st2.txt

Table AT7: List of miRNAs significantly correlated with bacteria previously associated with CRC. Each sheet includes the miRNAs correlated with one of the seven taxa (Fusobacterium, Providencia, Bacteroides, Akkermansia, Roseburia, Porphyromonas, and Peptostreptococcus) and provides the miRNA's identity, correlation coefficient, and pseudo-P value. Accessed through: https://journals.asm.org/doi/suppl/10.1128/mSystems.00205-17/suppl_file/sys003182230st3.xlsx

Table AT8: Predicted functions of miRNAs correlated with CRC-associated bacteria. Each sheet represents the predicted functions of miRNAs correlated with one of the taxa listed in the sheet's name. Accessed through: https://journals.asm.org/doi/suppl/10.1128/mSystems.00205-17/suppl_file/sys003182230st4.x00000000000000lsx